Foregone opportunities and choosing not to act: Replications of Inaction Inertia effect

*Jieying Chen
Department of Business Administration
University of Manitoba
jieying.chen@umanitoba.ca

*Long Sang Hui , *Teresa Yu
Department of Psychology, University of Hong Kong, Hong Kong SAR
yuteresa@connect.hku.hk, sam510@connect.hku.hk

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

**Shiyuan Vanessa Zeng (shiyuanv@connect.hku.hk), **Tze Lam Ching (luluctl@connect.hku.hk), **Chi Ho Wallace Ng (wallyng@connect.hku.hk), **Kwok Wai Frankie Wu (u3508971@connect.hku.hk), **Chung Man Yuen (jyycm@connect.hku.hk), **Tsz Ki Lau (jeanlau@connect.hku.hk)
Department of Psychology, University of Hong Kong, Hong Kong SAR

Bo Ley Cheng (boleyc@hku.hk), Ka Wai Ng (kawaing@hku.hk)
Department of Psychology, University of Hong Kong, Hong Kong SAR

In press at Social Psychological and Personality Science

Accepted for publication on Dec 20, 2019

*Contributed equally, joint first author
** Contributed equally, joint fifth author
^Corresponding author
Word: abstract – 150, manuscript - 3874

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 at the Department of Psychology, University of Hong Kong. His research focuses on morality, personal values, lay-beliefs, and judgment decision-making.

Jieying Chen is an assistant professor at the Department of Business Administration, University of Manitoba. Her research focuses on cross-cultural interactions, strategic human resource management, and mindfulness.

Teresa Yu, Long Sang Hui, Shiyuan Vanessa Zeng, Tze Lam Ching, Chi Ho Wallace Ng, Kwok Wai Frankie Wu, Chung Man Yuen, and Tsz Ki Lau are undergraduate students at the University of Hong Kong.

Cheng Bo Ley and Ka Wai Ng are teaching assistants at the Department of Psychology, University of Hong Kong.

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.

Acknowledgments or Author Notes: [None]

Authorship declaration:

Gilad was the course instructor for fundamentals and advanced social psychology courses (PSYC2020/3052) and led the two reported replication efforts in those courses. Gilad supervised each step in the project, conducted the pre-registrations, and ran data collection. Jieying integrated the two replication efforts with validation and extensions of the analyses. Jieying and Gilad jointly wrote the final manuscript.

Teresa Yu and Long Sang Hui conducted the U.S. replication as part of the advanced social psychology course (PSYC3052).

Shiyuan Vanessa Zeng, Tze Lam Ching, Chi Ho Wallace Ng, Kwok Wai Frankie Wu, Chung Man Yuen, and Tsz Ki Lau conducted the Hong Kong replication as part of the fundamentals of social psychology course (PSYC2020).
Cheng Bo Ley guided and assisted the replication effort in the PSYC3052 course. Ka Wai Ng guided and assisted the replication effort in the PSYC2020 course.

Abstract

Inaction inertia is the phenomenon that forgoing an initial attractive opportunity decreases the likelihood of taking a subsequent opportunity that is less attractive, even when the subsequent opportunity still offers positive value. We conducted three pre-registered replications of Tykocinski, Pittman, and Tuttle (1995) Experiments 1 and 2's four scenarios in four samples (N = 1555). We found consistent findings across samples, with the inaction inertia effect dependent on the scenario used. Strongest support was for the car scenario (*d* = -0.57 to -0.68) and the ski scenario (*d* = -0.18 to -0.67), with mixed findings for the fitness scenario (large-small: *d* = -0.62; control contrasts: opposite to predictions) and weak to no effects for the flyer scenario (*d* = -0.14 to 0.02). We conclude that context is important in studying inaction inertia, recommend the car and ski scenarios for follow-up research on inaction inertia, and discuss implications for future research.

Keywords: inaction inertia; decision making; pre-registered replication

Foregone opportunities and choosing not to act: Replications of Inaction Inertia effect

Inaction inertia is the phenomenon that forgoing an initial attractive opportunity decreases the likelihood of undertaking take a subsequent, less attractive opportunity (Tykocinski, Pittman, & Tuttle, 1995). This phenomenon occurs even when subsequent opportunity offers desired positive absolute value (Tykocinski & Pittman, 1998). Past experience serves as a reference point to which current decisions are compared: an initial decision to not act increases the likelihood of subsequent decisions to not act, especially if the subsequent opportunity fails to meet the value of the foregone opportunity.

Inaction inertia effect has been used to explain many other phenomena such as product change aversion (Tykocinski & Pittman, 2001), procrastination (Pittman, Tykocinski, Sandman-Keinan, & Matthews, 2007), and suboptimal decisions regarding retirement savings (Krijnen, Zeelenberg, Breugelmans, & van Putten, 2019), negotiations in job offers (Forster & Diab, 2017), and international affairs (Terris & Tykocinski, 2016). As an example from education, students who missed an initial opportunity to submit an assignment early for a big bonus were more likely to end up never submitting that assignment at all (e.g., Pittman et al., 2007). Another example is that when retirement plans change, and foregone plan opportunity was much more attractive (vs. slightly more attractive), people were less willing to save under the new retirement plan (Krijnen et al., 2019). Such behaviors have been explained by individuals' decreased subjective value of the subsequent opportunity and the desire to avoid regret over having missed the initial opportunity (Arkes, Kung, & Hutzel, 2002; Tykocinski & Pittman, 2001).

The chosen study for replication

We chose Tykocinski et al. (1995) for replication for several reasons. First, the article is considered the first comprehensive demonstration of the inaction inertia effect, resulting in an impactful area of inquiry in fields such as decision making, consumer behavior, motivation, emotion, behavioral economics, organizational management, and international politics (e.g., Anderson, 2003; Tykocinski & Pittman, 1998). Second, to our knowledge and based on communications with the authors, there have been no direct replications of the experiments conducted in the article. We aimed to address increasing calls for more pre-registered replication work and for promoting more open transparent reporting to increase credibility and trustworthiness of published findings (Gelman & Loken, 2013; Munafò et al., 2017; Nosek & Lakens, 2014), following recent mass pre-registered replication attempts that found low replicability rates of classic findings in the field (Camerer et al., 2018; Open Science Collaboration, 2015; Klein et al., 2018).

Tykocinski et al. (1995) showed the inaction inertia effect by demonstrating that the gap between the initial opportunity and the subsequent opportunity affected the choice made for the subsequent opportunity. The authors compared three conditions manipulating differences between the initial and the subsequent opportunity: large-difference, small-difference, and control (with no initial opportunity). They found that the more attractive a foregone initial opportunity was compared to the subsequent opportunity, the lower the likelihood of taking action on the subsequent opportunity.

In the first two experiments, Tykocinski et al. (1995) tested the inaction inertia effect in four scenarios: ski, car, frequent flyer, and fitness center. These scenarios involved either money or effort, with differences in numerical values. The authors found support for the inaction inertia effect in all four scenarios, yet pairwise comparisons showed nuanced differences across scenarios. For example, the difference between the large-difference condition and the control condition was significant in the ski scenario, but non-significant in the fitness center scenario. Also, the difference between the large-difference condition and the small-difference condition was significant in the fitness center scenario, but non-significant in the car scenario. In their Experiments 3 to 6, the authors examined possible explanations for inaction inertia, including perceptual contrast, dissonance, self-perception, and commitment, which yielded mixed findings. In these experiments, only the large-difference condition and the small-difference condition were included. We chose the simpler first two experiments that focused on the main effect and included all three conditions for replication. A summary of the original findings in the target article Experiments 1 and 2 is shown in Table 1.

Table 1

Summary of findings in Tykocinski et al. (1995) Experiments 1 and 2

	Condition				
Scenario	Large-difference	Small-difference	Control	*F*	*p*
Ski	4.94_b	7.25_a	6.36_a	*F*(2,105) = 5.92	.003
Car	3.39_b	$4.12_{a,b}$	5.05_a	*F*(2,117) = 4.12	.02
Frequent flyer	5.63_b	7.45_a	7.36_a	*F*(2,117) = 5.46	.006
Fitness	5.34_b	7.26_a	$6.26_{a,b}$	*F*(2,117) = 3.92	.03
Across car, frequent flyer, and fitness	4.72_b	6.25_a	6.29_a	*F*(2,117) = 10.93	.0002

Note. Summary adapted from Tykocinski et al. (1995) Table 1 (p. 796). Subscripts indicate whether there was a difference found between means at $p < .05$ (Tukey's). The ski scenario was tested in Experiment 1, and the car, frequent flyer, and fitness scenarios were tested in Experiment 2.

Overview of replications

We conducted three pre-registered replications of Tykocinski et al.'s (1995). Experiment 1 was conducted with an undergraduate student sample in Hong Kong, which served as an initial examination of the inaction inertia effect and its effect size. Experiment 2 was conducted with American Amazon Mechanical Turk (MTurk) online labour market workers. In this experiment, the sample size was larger, and participants are diverse in terms of age, ethnicity, and educational level. Experiment 3 was conducted with two groups of MTurk workers, which allowed us to compare results obtained using a between-subject (different conditions) design versus a mixed factorial (same condition) design. In all experiments, we combined the four scenarios used in Tykocinski et al.'s (1995) Experiments 1 and 2 into a single experiment with randomized order contrasting the three conditions: large-difference, small-difference, and control.

Method

Pre-registration and open-science

In all replications, we first pre-registered the experiment on the Open Science Framework and data collection was launched soon after. Pre-registrations, power analyses, disclosures, and all materials used in the experiments are available in the supplementary. These together with data and code were shared on the Open Science Framework (pre-registrations, datasets, and code: https://osf.io/kxe73/; Experiment 1 pre-registration: https://osf.io/kbnjw; Experiment 2 pre-registration: https://osf.io/e93k4; Experiment 3 pre-registration: https://osf.io/83w2x).

Power analyses and exclusions

We used Lakens' (2013) effect size calculator and DeCoster's (2012) effect size converter to determine the effect size of one-way ANOVA in the target article. The Cohen's *f* effects of the scenarios were: ski - 0.34, car - 0.27, frequent flyer - 0.31, and fitness center - 0.26. The Cohen's *f* of the one-way ANOVA across the four scenarios was 0.43. To obtain power of 95% at an alpha of 5%, for the minimum effect of .26, the required sample size was 234. The target article did not contain the information required to calculate the effect sizes of pairwise comparisons.

The number of participants for Experiment 1 was limited to the students who took the course and conducted replications of classic findings in judgment and decision making (see details below). The sample in Experiment 1 was knowingly under-powered. We, therefore, caution using Null Hypothesis Significance Testing findings (*p*-values) in interpreting Experiment 1 results, and suggest that instead readers focus on effect sizes, interpret them in comparison to those found in the well-powered Experiments 2 and 3, and evaluate replication findings overall using the mini meta-analytic summary provided at the end.

In the target article, the authors did not mention any exclusion practice, we therefore assumed that no participants were excluded. In this replication analysis, we followed the same rule and did not further exclude any participants. We also conducted a set of supplementary analyses based on pre-specified exclusion criteria (e.g., understand the English used in this study, have seen the materials, were serious when completing the study, correctly guessed the hypothesis) that were proposed in the pre-registration. Overall, we found that exclusions had little to no impact on the results, and that participants were generally proficient in English and serious. See supplementary materials for the results after applying these exclusion criteria.

Participants

Experiment 1 - Hong Kong undergraduate sample

The first replication was considered a pre-test in an undergraduate course at a university in Hong Kong. Students were randomly assigned to work in groups of 3 to 6 people, with each group responsible for conducting one replication, and one of the groups was in charge of the current replication. Students then served as the target sample for the experiments designed by their classmates, which they were not familiar with. The course materials covered classic judgment and decision-making literature, which meant that students were made aware of a wide array of heuristics and biases. The experiment can therefore be considered as a very conservative test of the effect in a non-naive sample. The final sample included 43 student participants (13 men, 30 women; M_{age} = 20.20, SD_{age} = 1.00).

Experiments 2 and 3 - online American Amazon Mechanical Turk

Participants in Experiments 2 and 3 were recruited online from MTurk using Turkprime.com platform (Litman, Robinson, & Abberbock, 2017). In Experiment 2, we recruited a total of 309 participants (140 men, 169 women; M_{age} = 38.42, SD_{age} = 11.53).

In Experiment 3, we recruited a total of 1203 participants. Participants who took part in Experiment 2 were not allowed to take part in Experiment 3. We randomly assigned participants into one of the two designs: different conditions design (603 participants; 263 men, 340 women, M_{age} = 40.40, SD_{age} = 12.25) and same condition design (600 participants; 299 men, 301 women, M_{age} = 40.40, SD_{age} = 12.19). In the "different conditions" design, participants were randomly assigned to one of the conditions (i.e., large-difference, small-difference, control) for each of the scenarios, thus they could be assigned to different conditions in different scenarios. This was also the design used in Experiments 1 and 2. In the "same condition" design, participants were consistently assigned to the same condition across all scenarios, to allow for a mixed factorial design.

Procedures and measures

The experimental materials were identical to those used in Experiments 1 and 2 in Tykocinski et al. (1995). Their Experiment 1 introduced the ski scenario, and their Experiment 2 tested three scenarios: car, frequent flyer, and fitness center. The presentation order in

Experiment 2 did not have any effect. Because the four scenarios have similar structure and design, we combined the four scenarios into a single experiment, and randomized the presentation order of the four scenarios.

In our replication studies Experiment 1, Experiment 2, and Experiment 3 (different conditions design), in each scenario, participants were randomly assigned to one of three conditions in a between-subjects design: large-difference, small-difference, and control conditions. In Experiment 3 (same condition design), participants were consistently assigned to the same condition in all scenarios, resulting in a 4 (scenario) x 3 (condition) mixed factorial design, where scenario was a within-person factor, and condition was a between-person factor.

All conditions first described a foregone opportunity. In the large-difference condition, the subsequent opportunity was much less attractive compared to the foregone opportunity (e.g., $40 compared to $90). In the small-difference condition, the subsequent opportunity was slightly less attractive compared to the foregone opportunity (e.g., $80 compared to $90). In the control condition, only the second opportunity was presented.

After reading each scenario, participants indicated the likelihood that they would take action on the second opportunity on an 11-point Likert scale (0 = *not at all likely*, 10 = *extremely likely*), which was the dependent variable of the study.

A summary of the experimental materials is shown in Table 2. The survey ended with a funneling section and collection of general demographic information.

Analytical approach

For Experiment 1, Experiment 2, and Experiment 3 (different conditions design) we conducted one-way ANOVAs to examine whether there was a difference in the likelihood to act on the subsequent opportunity among these conditions. For Experiment 3 (same condition design), we conducted a mixed factorial ANOVA. When there was a significant main effect or interaction effect, we performed pairwise comparisons to identify which pair of conditions differ from one another. All *p* values were adjusted using the appropriate method based on the tests of statistical assumptions. Confidence intervals of Cohen's *d* and partial or generalized eta-squared were computed using package apaTables (Stanley, 2018) and MBESS (Kelley, 2007) in R (R Core Team, 2019).

In both designs, a participant read four scenarios, thus responses in multiple scenarios might not be fully independent. We did not know whether any potential spillover effect across scenarios will affect the results of between-person comparisons across conditions. We therefore also analyzed the data of Experiment 1, Experiment 2, and Experiment 3 (Different Conditions Design) using mixed effects models with package lme4 (Bates, Maechler, Bolker, & Walker, 2015) in R, and the results were available in supplementary materials. For Experiment 3 (Same Condition Design), we also conducted one-way ANOVAs using condition as the between-person factor and reported the results in supplementary materials. The results were largely consistent regardless of the design or the method of analysis.

We conducted a series of mini meta-analyses of the effect sizes of the large vs. small comparison, large vs. control comparison, and small vs. control comparison. The results were visualized using forest plots with package metaviz (Kossmeier, Tran, & Voracek, 2019) in R.

Table 2

Summary of the scenarios and measures

Scenario	Large-Difference Condition	Small-Difference Condition	Control Condition	Dependent Variable
Ski Resort	Your friend called you at the beginning of October and told you that he intended to buy a special pass to Ski Liberty (where you both like to ski). He said that the deal was that if you bought the pass before the 15th of October the pass would cost you only **$40** instead of the **$100** regular price. Although it sounded like a good idea, you forgot to do it by the 15th. The next time your friend called he told you that although you missed the deadline you could still get a pass for **$90** if you pay this week.	Your friend called you at the beginning of October and told you that he intended to buy a special pass to Ski Liberty (where you both like to ski). He said that the deal was that if you bought the pass before the 15th of October the pass would cost you only **$80** instead of the **$100** regular price. Although it sounded like a good idea, you forgot to do it by the 15th. The next time your friend called he told you that although you missed the deadline you could still get a pass for **$90** if you pay this week.	Your friend called you at the beginning of October and told you that he intended to buy a special pass to Ski Liberty (where you both like to ski). He said that the deal was that if you buy the pass this week, it will only cost you **$90** instead of the **$100** regular price.	How likely are you to spend the $90? (0 = *not at all likely*; 10 = *extremely likely*)
Car	You see a television advertisement by a local dealer for a car and you are interested in buying. The advertisement promotes a limited-time **$500** factory rebate on the car, providing it is purchased this week. However, you had seen this advertisement once before. Back then, the car was offered with a larger rebate value, **$2,500**, for a limited time. Although you were interested in the deal at that time, you had missed the deadline.	You see a television advertisement by a local dealer for a car and you are interested in buying. The advertisement promotes a limited-time **$500** factory rebate on the car, providing it is purchased this week. However, you had seen this advertisement once before. Back then, the car was offered with a larger rebate value, **$750**, for a limited time. Although you were interested in the deal at that time, you had missed the deadline.	You see a television advertisement by a local dealer for a car and you are interested in buying. The advertisement promotes a limited-time **$500** factory rebate on the car, providing it is purchased this week.	How likely are you to buy the car this week? (0 = *not at all likely*; 10 = *extremely likely*)
Frequent Flyer	You are considering joining a frequent flyer program before taking a trip for the holidays. If you decide to join, you would accumulate **5,500** miles (towards a free ticket, on accumulating 20,000 miles). You had considered joining this program once before, near the beginning of the year, but had not done so. The number of miles that would have been accumulated after this current trip had you joined earlier was **15,500** miles.	You are considering joining a frequent flyer program before taking a trip for the holidays. If you decide to join, you would accumulate **5,500** miles (towards a free ticket, on accumulating 20,000 miles). You had considered joining this program once before, near the beginning of the year, but had not done so. The number of miles that would have been accumulated after this current trip had you joined earlier was **7,500** miles	You are considering joining a frequent flyer program before taking a trip for the holidays. If you decide to join, you would accumulate **5,500** miles (towards a free ticket, on accumulating 20,000 miles).	How likely are you to join the program? (0 = *not at all likely*; 10 = *extremely likely*)
Fitness Center	You are considering becoming a member of a fitness club located **30 min** away. You could have joined another center closer to home, but, having neglected to act quickly, had missed the opportunity. The center that by now had closed its membership roll was said to be located **5 min** away from your home.	You are considering becoming a member of a fitness club located **30 min** away. You could have joined another center closer to home, but, having neglected to act quickly, had missed the opportunity. The center that by now had closed its membership roll was said to be located **25 min** away from your home.	You are considering becoming a member of a fitness club located **30 min** away.	How likely are you to become a member of the fitness club now? (0 = *not at all likely*; 10 = *extremely likely*)

Note. The ski resort scenario came from Experiment 1 of Tykocinski et al. (1995), and the other three scenarios came from Experiment 2 of Tykocinski et al. (1995). All materials were identical to those used in the target article.

Replications evaluation

We evaluted the three replications in terms of methodological similarity to the target article based on the taxonomy introduced by LeBel, McCarthy, Earp, Elson, and Vanpaemel (2018). This taxonomy categorizes replications into a continuum of exact replication to very far replication using a set of key design facets, including operationalizations of independent variables and dependent variables, stimuli used to measure independent variables and dependent variables, procedural details, physical settings, and contextual variables. Adopting this taxonomy, the three experiments adhered to the criteria of very close replications. All design aspects were the same as in the target article except for procedural details, physical setting, and contextual variables. A summary of the comparisons on key design facets is provided in supplementary materials.

Results

Descriptive statistics of all conditions of all scenarios in the three experiments are provided in Table 3. Violin jitter dot plots (Patil, 2018) of the data are provided in the supplementary.

Table 3
Descriptive Statistics

Experiment 1 Hong Kong Student Sample	Scenario 1 Ski			Scenario 2 Car			Scenario 3 Frequent Flyer			Scenario 4 Fitness Center		
Condition	*N*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Large-difference	13	6.00	2.83	15	2.93	2.58	14	6.86	2.88	15	4.80	2.73
Small-difference	14	6.71	2.87	14	6.64	1.74	14	6.86	2.71	15	6.67	3.56
Control	16	6.69	2.24	14	6.07	2.06	15	8.40	1.12	13	3.31	1.55

Experiment 2 American MTurk Sample	Scenario 1 Ski			Scenario 2 Car			Scenario 3 Frequent Flyer			Scenario 4 Fitness Center		
Condition	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Large-difference	106	4.39	3.08	104	2.52	2.31	104	5.88	2.98	101	3.47	2.69
Small-difference	101	6.84	3.11	101	4.50	2.78	104	6.27	2.66	104	5.11	2.83
Control	102	5.67	2.97	104	4.25	2.92	101	6.74	2.71	104	2.41	2.87

Experiment 3 MTurk (Different Conditions)	Scenario 1 Ski			Scenario 2 Car			Scenario 3 Frequent Flyer			Scenario 4 Fitness Center		
Condition	*N*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Large-difference	201	5.29	2.87	201	2.83	2.66	201	6.58	2.90	201	3.28	2.86
Small-difference	201	6.98	2.79	201	4.12	3.00	201	5.96	2.76	201	5.01	3.25
Control	201	5.79	3.16	201	4.15	2.67	201	6.72	2.86	201	2.19	2.50

Experiment 3 MTurk (Same Condition)	Scenario 1 Ski			Scenario 2 Car			Scenario 3 Frequent Flyer			Scenario 4 Fitness Center		
Condition	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Large-difference	199	5.09	3.01	199	2.76	2.61	199	6.20	2.83	199	3.14	2.86
Small-difference	200	7.13	2.54	200	4.49	2.87	200	6.38	2.52	200	5.13	3.25
Control	201	5.08	3.17	201	3.99	2.97	201	6.34	3.05	201	2.04	2.60

For Experiment 1, Experiment 2, and Experiment 3 (different conditions design), we performed one-way analysis of variance (ANOVA) and two-tail pairwise comparisons on the mean likelihood of taking action on the subsequent opportunity in the four scenarios (ski, car, frequent flyer, fitness center) contrasting the three conditions: large-difference, small-difference and control. Summaries are provided in Tables 4 and 5.

Table 4
One-way ANOVA on the Mean Likelihood of Acting on the Subsequent Opportunity

Experiment 1 (*n* = 43)	Conditions										
Scenario	Large-difference	Small-difference	Control	*df*	*F*	*p*	*f*	η^2	90% CI for η^2	95% CI for η^2
Ski	6.00	6.71	6.69	2, 40	0.32	.73	.1	.02	[.00, .08]	[.00, .12]
Car	2.93[a]	6.64[b]	6.07[b]	2, 40	12.49	< .00	.7	.38	[.17, .52]	[.13, .54]
Frequent flyer	6.86	6.86	8.40	2, 40	2.10	.14	.3	.10	[.00, .23]	[.00, .26]
Fitness center	4.80[a,b]	6.67[a]	3.31[b]	2, 40	5.12	.01	.5	.20	[.03, .35]	[.01, .38]
Experiment 2 (*n* = 309)	**Conditions**									
Scenario	Large-difference	Small-difference	Control	*df*	*F*	*p*	*f*	η^2	90% CI for η^2	95% CI for η^2
Ski	4.39[a]	6.84[b]	5.67[c]	2, 600	16.76	< .00	.3	.10	[.05, .15]	[.04, .16]
Car	2.52[a]	4.50[b]	4.25[b]	2, 600	16.64	< .00	.3	.10	[.05, .15]	[.04, .16]
Frequent flyer	5.88	6.27	6.74	2, 600	2.49	.08	.1	.02	[.00, .04]	[.00, .05]
Fitness center	3.47[a]	5.11[b]	2.41[c]	2, 600	24.42	< .00	.4	.14	[.08, .19]	[.07, .21]
Experiment 3 (Different Conditions) (*n* = 603)	**Conditions**									
	Large-difference	Small-difference	Control	*df*	*F*	*p*	*f*	η^2	90% CI for η^2	95% CI for η^2
Ski	5.29[a]	6.98[b]	5.79[a]	2, 600	17.34	< .00	.2	.06	[.03, .08]	[.02, .09]
Car	2.83[a]	4.12[b]	4.15[b]	2, 600	14.94	< .00	.2	.05	[.02, .08]	[.02, .08]
Frequent flyer	6.58[a,b]	5.96[a]	6.72[b]	2, 600	4.03	.02	.1	.01	[.00, .03]	[.00, .03]
Fitness center	3.28[a]	5.01[b]	2.19[c]	2, 600	48.69	< .00	.4	.14	[.10, .18]	[.09, .19]

Note. For each scenario where the *F* test was significant, means of subgroups that do not share a common subscript are significantly different at $p < .05$. Depending on tests of statistical assumptions, we chose the appropriate method to conduct pairwise comparisons (see test results and details in supplementary). In Experiment 1, we conducted Hochberg's test for Scenarios 1, 2, and 3, and Games-Howell test for Scenario 4. In Experiment 2, we conducted Tukey's test for Scenarios 1, 3, and 4, and Games-Howell test for Scenario 2. In Experiment 3 (Different Conditions), we conducted Tukey's test for between-subject design Scenario 3, and the Games-Howell test for all other scenarios. Higher numbers indicate higher likelihoods of taking action on a subsequent opportunity on an 11-point scale.

Table 5
Pairwise Comparisons of the Likelihood of Acting on Subsequent Opportunity

Experiment 1 (*n* = 43)	Mean Difference	*t*	*p*	Cohen's *d*	95% CI for Cohen's *d*
Ski					
large vs. small	-0.71	-0.70	.86	-0.27	[-1.03, 0.49]
large vs. control	-0.69	-0.70	.86	-0.26	[-0.99, 0.48]
small vs. control	0.03	0.03	1.00	0.01	[-0.71, 0.73]
Car					
large vs. small	-3.70	-4.62	< .001	-1.72	[-2.56, -0.85]
large vs. control	-3.14	-3.90	.001	-1.45	[-2.26, -0.62]
small vs. control	0.57	0.70	.86	0.26	[-0.48, 1.01]
Frequent flyer					
large vs. small	0.00	0.00	1.00	0.00	[0-.74, 0.74]
large vs. control	-1.54	-1.77	.23	-0.66	[-1.40, 0.10]
small vs. control	-1.54	-1.77	.23	-0.66	[-1.40, 0.10]
Fitness center					
large vs. small	-1.87	-1.61	.26	-0.59	[-1.31, 0.15]
large vs. control	1.49	1.81	.19	0.69	[-0.09, 1.45]
small vs. control	3.36	3.31	.01	1.25	[0.43, 2.06]

Experiment 2 (*n* = 309)	Mean Difference	*t*	*p*	Cohen's *d*	95% CI for Cohen's *d*
Ski					
large vs. small	-2.45	-5.79	< .001	-0.81	[-1.09, -0.52]
large vs. control	-1.28	-3.02	.01	-0.42	[-0.69, -0.14]
small vs. control	1.17	2.74	.02	0.38	[0.11, 0.66]
Car					
large vs. small	-1.98	-5.27	< .001	-0.74	[-1.02, -0.45]
large vs. control	-1.73	-4.65	< .001	-0.64	[-0.92, -0.37]
small vs. control	0.25	0.65	.81	0.09	[-0.18, 0.36]
Frequent flyer					
large vs. small	-0.39	-1.02	.56	-0.14	[-0.41, 0.13]
large vs. control	-0.87	-2.23	.07	-0.31	[-0.59, -0.04]
small vs. control	-0.47	-1.22	.44	-0.17	[-0.44, 0.10]
Fitness center					
large vs. small	-1.64	-4.19	< .001	-0.59	[-0.86, -0.30]
large vs. control	1.05	2.69	.02	0.38	[0.10, 0.65]
small vs. control	2.69	6.93	< .001	0.96	[0.67, 1.25]

Table 5 (Continued)

Pairwise Comparisons of the Likelihood of Acting on Subsequent Opportunity

Experiment 3 (Different Conditions) (*n* = 603)	Mean Difference	*t*	*p*	Cohen's *d*	95% CI for Cohen's *d*
Ski					
large vs. small	-1.68	-5.96	< .001	-0.59	[-0.79, -0.39]
large vs. control	-0.49	-1.64	.23	-0.16	[-0.36, 0.03]
small vs. control	1.19	4.00	< .001	0.40	[0.20, 0.60]
Car					
large vs. small	-1.30	-4.59	< .001	-0.46	[-0.66, -0.26]
large vs. control	-1.33	-4.99	< .001	-0.50	[-0.70, -0.30]
small vs. control	-.03	-.11	.99	-0.01	[-0.21, 0.18]
Frequent flyer					
large vs. small	0.62	2.17	.08	0.22	[0.02, 0.41]
large vs. control	-0.14	-0.49	.88	-0.05	[-0.24, 0.15]
small vs. control	-0.76	-2.67	.02	-0.27	[-0.46, -0.07]
Fitness center					
large vs. small	-1.73	-5.65	< .001	-0.56	[-0.76, -0.36]
large vs. control	1.09	4.07	< .001	0.41	[0.21, 0.60]
small vs. control	2.82	9.74	< .001	0.97	[0.76, 1.18]

Note. Experiment 1: Hochberg's test for Scenarios 1, 2, and 3, and Games-Howell test for Scenario 4. Experiment 2, Tukey's test for Scenarios 1, 3, and 4, and Games-Howell test for Scenario 2. In Experiment 3 (Different Conditions), we conducted Tukey's test for between-subject design Scenario 3, and Games-Howell test for all other scenarios. Higher numbers indicate higher likelihoods of taking action on a subsequent opportunity on an 11-point scale.

For Experiment 3 (same condition design) we performed a mixed factorial ANOVA on the likelihood of acting on the subsequent opportunity (Table 6). The within-person factor was the scenario, and the between-person factor was the condition. Despite that Mauchly's W test was significant ($p = .01$), the Greenhouse-Geisser epsilon was .98, which was close to 1 and much higher than the lower-bound of .33, suggesting that sphericity was not a concern. The scenario main effect was significant (F (3, 1791) = 200.12, $p < .001$), showing support for the likelihood of acting on the subsequent opportunity differing across scenarios. The main effect of condition was also significant (F (2, 597) = 39.01, $p < .001$), showing support for the likelihood of acting on the subsequent opportunity differing across conditions. Pairwise comparisons (Table 7) suggested that the likelihood of acting was lower in the large-difference condition than in the small-difference condition (*d* = -0.78, 95% CI = [-0.99, -0.58], *t* = -7.81, *p* < .001); higher in the small-difference condition than in control condition (*d* = 0.75, 95% CI = [0.54, 0.95], *t* = 7.47, *p* < .001). There was no significant difference between the large-difference condition and the control condition (*d* = -0.04, 95% CI = [-0.23, 0.16], *t* = -0.36, *p* = .93).

The interaction between scenario and condition was significant (*F* (6, 1791) = 19.83, *p* < .001). The results of multiple pairwise comparisons are summarized in Table 8. We also conducted one-way ANOVA using condition as a between-subject factor for each of the scenarios in this sample (see supplementary materials), and the results were similar to those reported here.

Table 6

Results of Mixed Factorial ANOVA for Experiment 3 (Same Condition Design) (n = 600)

	Sum of Squares	*df*	*F*	*p*	f_G	f_p	η^2_G	η^2_p	90% CI for η^2_p	95% CI for η^2_p
(Within-subject)										
Scenario	3704.53	3	200.12	< .001	.43	.58	.16	.25	[.22, .28]	[.22, .28]
Scenario * Condition	734.00	6	19.83	< .001	.19	.25	.04	.06	[.04, .08]	[.04, .08]
$\text{Residual}_{(\text{Within})}$	11051.32	1791								
(Between-subject)										
Condition	1122.21	2	39.01	< .001	.24	.37	.05	.12	[.08, .15]	[.07, .16]
$\text{Residual}_{(\text{Between})}$	8586.78	597								

Table 7

Pairwise Comparisons of Mixed Factorial ANOVA for Experiment 3 (same condition design) (N = 600)

	Mean Difference	*t*	P_{Tukey}	Cohen's *d*	95% CI for Cohen's *d*
Across scenarios					
large vs. small	-1.48	-7.81	< .001	-0.78	[-0.99, -0.58]
large vs. control	-0.07	-0.36	.93	-0.04	[-0.23, 0.16]
small vs. control	1.42	7.47	< .001	0.75	[0.54, 0.95]
Ski					
large vs. small	-2.04	-7.10	< .001	-0.71	[-0.91, -0.51]
large vs. control	0.01	0.02	1.00	0.00	[-0.19, 0.20]
small vs. control	2.05	7.14	< .001	0.71	[0.51, 0.91]
Car					
large vs. small	-1.73	-6.03	< .001	-0.60	[-0.80, -0.40]
large vs. control	-1.23	-4.29	.001	-0.43	[-0.63, -0.23]
small vs. control	0.50	1.75	.85	0.17	[-0.02, 0.37]
Frequent flyer					
large vs. small	-0.18	-0.62	1.00	-0.06	[-0.26, 0.13]
large vs. control	-0.15	-0.51	1.00	-0.05	[-0.25, 0.15]
small vs. control	0.03	0.11	1.00	0.01	[-0.18, 0.21]
Fitness center					
large vs. small	-1.98	-6.91	< .001	-0.69	[-0.89, -0.49]
large vs. control	1.10	3.84	.01	0.38	[0.19, 0.58]
small vs. control	3.09	10.77	< .001	1.08	[0.87, 1.28]

Note. The *p* values of pairwise comparisons were adjusted using the Tukey's honestly significant difference method.

A comparison of findings in the target article and the three pre-registered replications is provided in Table 8. Forest plots of the effect sizes and confidence intervals for the pairwise effects and mini-meta summaries are provided in Figures 1 (large versus small), 2 (large versus control), and 3 (small versus control). The mini meta-analyses were meant to provide an overall estimation of effect sizes and probabilities when there were multiple studies with both significant and nonsignificant findings (Goh, Hall, & Rosenthal, 2016; Lakens & Etz, 2017). These analyses also met the requirements of credible internal meta-analyses, as we included all studies conducted and attempted planned, pre-registered tests only (Vosgerau, Nelson, Simonsohn, & Simmons, 2019).

Findings across the three replications were largely consistent, with similar effects found in the four samples. However, findings differed across scenarios. Findings in the ski scenario and car scenario were clearest and in line with the inaction inertia effect, suggesting largely successful replications. In contrast, in all but one sample (Experiment 3 different conditions design), we found no support for differences among conditions in the frequent flyer scenario, indicating a likely failed replication for frequent flyer scenario. In all samples, there was a significant difference among conditions in the fitness center scenario, although the direction of the large-difference condition versus control condition effects was not in the expected direction.

Likelihood to act on the subsequent opportunity was lower for the large difference condition than for the small difference condition (mini meta ski: d = -0.67, 95% CI [-0.80, -0.55]; car: d = -0.68, CI [-0.93, -0.42]; fitness center: d = -0.62, CI [-0.74, -0.49]; overall: d = -0.49, CI [-0.67, -0.32]).

For the large-difference vs. control comparison, there were mixed findings: in the car scenario, the likelihood to act was lower in the large-difference condition than in the control condition (mini meta car: d = -0.57, CI [-0.77, -0.36]), whereas in the fitness center condition, the likelihood to act was higher in the large-difference condition than in the control condition (ski: d = 0.40, CI [0.28, 0.52]). The effect sizes in the ski scenario and the frequent flyer scenario were small, and their confidence intervals included zero. The overall effect size of the large-difference vs. control comparison was d = -0.15, CI [-0.35, 0.04].

The contrast between the small and control conditions revealed a positive (mini-meta ski: d = 0.47, CI [0.25, 0.69]; fitness center: d = 1.02, CI [0.89, 1.15]) or null effect (car: d = 0.09, CI [-0.03, 0.21]; frequent flyer: d = -0.17, CI [-0.36, 0.02]). The overall effect size of the small-difference vs. control comparison was d = 0.33, CI [0.09, 0.57].

Table 8

Summary of Effect Comparisons among the Target Article and the Replication Studies

Scenario	Statistical test	Target Article (n = 108~120)	Replication Experiment 1 (n = 43)	Replication Experiment 2 (n = 309)	Replication Experiment 3 Different Conditions (n = 603)	Replication Experiment 3 Same Condition (n = 600)	Conclusion
Across Scenarios	Mixed ANOVA					f_G = .43	
Ski	One-way ANOVA	f = .34	f = .13	f = .33	f = .24		Mostly consistent findings across the four samples. Replicated: one-way ANOVA and large vs. small comparison (in all samples but Experiment 1). Deviations: large vs. control comparison was not significant in some samples; small vs. control comparison was significant in the *opposite direction* of prediction. Summary: **Mostly successful**.
	Large vs. Small	< 0	d = -.27 [-1.03, .49]	d = -.81 [-1.09, -.52]	d = -.59 [-.79, -.39]	d = -.71 [-.91, -.51]	
	Large vs. Control	< 0	d = -.26 [-.99, .48]	d = -.42 [-0.69, -.14]	d = -.16 [-.36, .03]	d = .00 [-.19, .20]	
	Small vs. Control	n.s.	d = .01 [-.71, .73]	d = .38 [.11, .66]	d = .40 [.20, .60]	d = .71 [.51, .91]	
Car	One-way ANOVA	f = 0.27	f = .79	f = .33	f = .22		Consistent findings across the four samples. Replicated: one-way ANOVA; large vs. control comparison. Deviations: The large vs. small comparison was non-significant in the target article and but significant in the replication study in the *expected direction*. Summary: **Mostly successful**.
	Large vs. Small	n.s.	d = -1.72 [-2.56, -.85]	d = -.74 [-1.02, -.45]	d = -.46 [-.66, -.26]	d = -.60 [-.80, -.40]	
	Large vs. Control	< 0	d = -1.45 [-2.26, -.62]	d = -.64 [-.92, -.37]	d = -.50 [-.70, -.30]	d = -.43 [-.63, -.23]	
	Small vs. Control	n.s.	d = .26 [-.48, 1.01]	d = .09 [-.18, .36]	d = -.01 [-.21, .18]	d = .17 [-.02, .37]	

Scenario	Statistical test	Target Article (*n* = 108~120)	Replication Experiment 1 (*n* = 43)	Replication Experiment 2 (*n* = 309)	Replication Experiment 3 Different Conditions (*n* = 603)	Replication Experiment 3 Same Condition (*n* = 600)	Conclusion
Frequent Flyer	One-way ANOVA	*f* = .31	*f* = .32	*f* = .13	*f* = .11		Mostly consistent findings across the four samples. Deviations: Very weak or null effects for the main effect or any pairwise comparison. Summary: **Likely failed replication**.
	Large vs. Small	< 0	*d* = .00 [-.74, .74]	*d* = -.14 [-.41, .13]	*d* = .22 [.02, .41]	*d* = -.06 [-.26, .13]	
	Large vs. Control	< 0	*d* = -.66 [-1.40, .10]	*d* = -.31 [-.59, -.04]	*d* = -.05 [-.24, .15]	*d* = -.05 [-.25, .15]	
	Small vs. Control	n.s.	*d* = -.66 [-1.40, .10]	*d* = -.17 [-.44, .10]	*d* = -.27 [-.46, -.07]	*d* = .01 [-.18, .21]	
Fitness Center	One-way ANOVA	*f* = .26	*f* = .51	*f* = .40	*f* = .40		Mostly consistent findings across the four samples. Replicated: one-way ANOVA, large vs. small comparison. Deviations: findings regarding control condition deviated and was positioned lower than the large differences condition. Therefore, all contrasts against the control were in the *opposite direction* to predictions. Summary: **Mostly successful for large versus small contrast**. *Control condition needs to be re-evaluated.*
	Large vs. Small	< 0	*d* = -.59 [-1.31, .15]	*d* = -.59 [-.86, -.30]	*d* = -.56 [-.76, -.36]	*d* = -.69 [-.89, -.49]	
	Large vs. Control	n.s.	*d* = .69 [-.09, 1.45]	*d* = .38 [.10, .65]	*d* = .41 [.21, .60]	*d* = .38 [.19, .58]	
	Small vs. Control	n.s.	*d* = 1.25 [.43, 2.06]	*d* = .96 [.67, 1.25]	*d* = .97 [.76, 1.18]	*d* = 1.08 [.87, 1.28]	

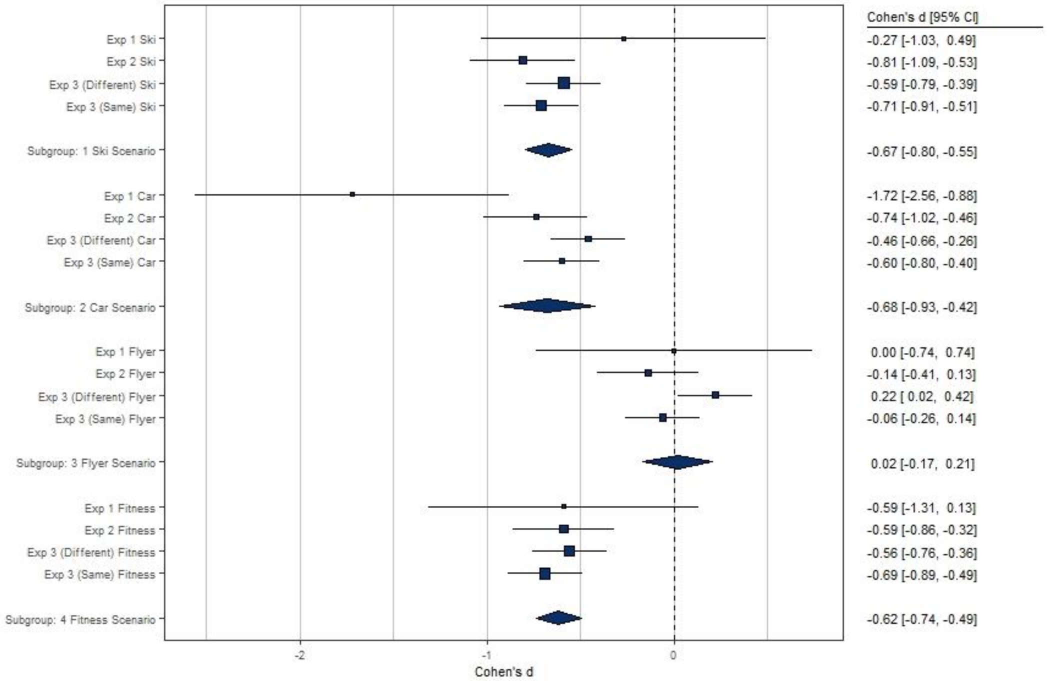


Figure 1. **Large-difference** versus **small**-difference conditions: Forest plot. Mini meta-analytic summaries were calculated using random effects DL model in the metaviz R module. The overall effect size was *d* = -0.49, CI [-0.67, -0.32].

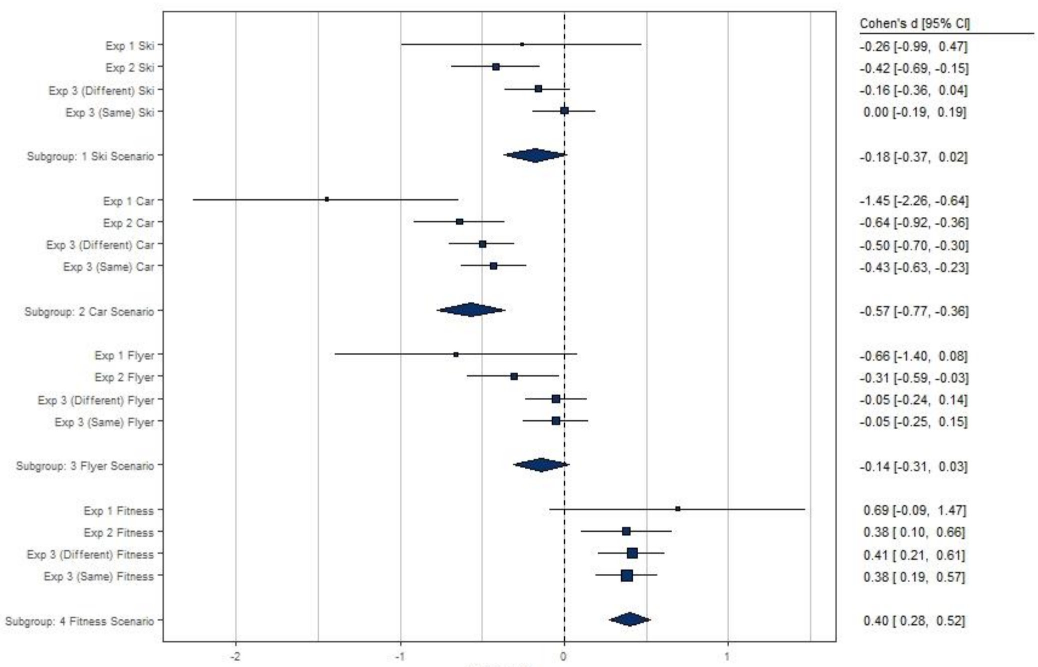


Figure 2. **Large-difference** versus **control** conditions: Forest plot. Mini meta-analytic summaries were calculated using random effects DL model in the metaviz R module. The overall effect size was *d* = -0.15, CI [-0.35, 0.04].

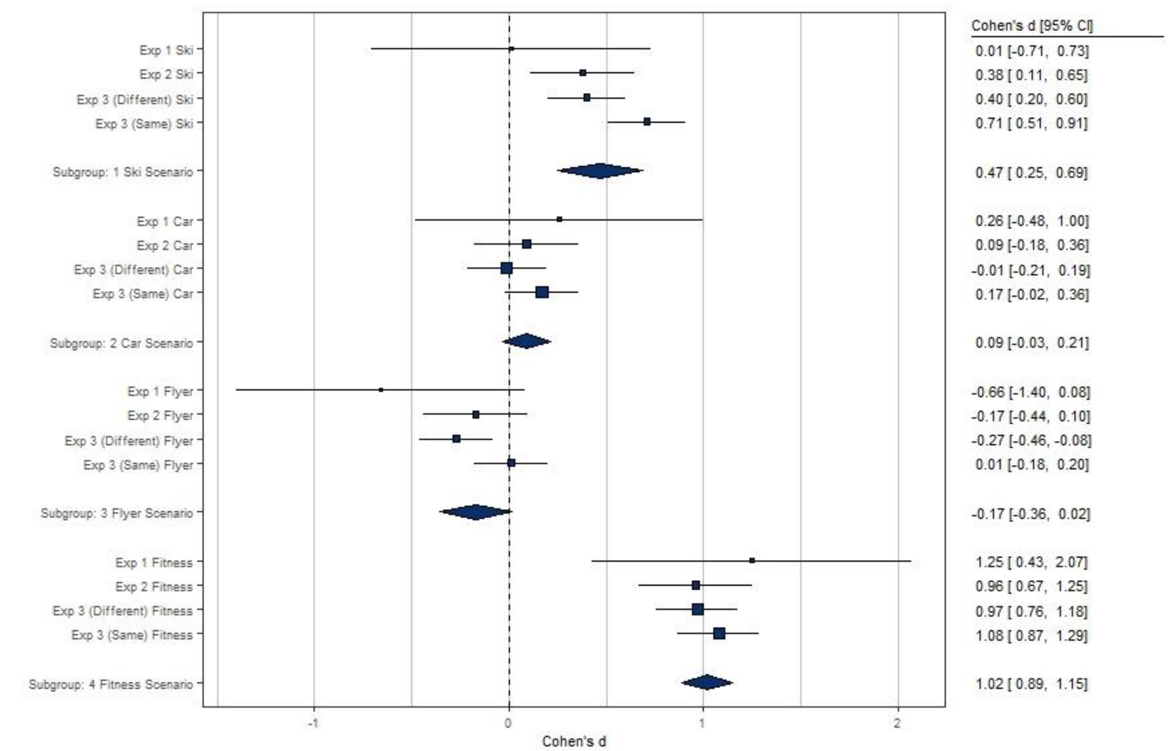


Figure 3. **Small-difference** versus **control** conditions: Forest plot. Mini meta-analytic summaries were calculated using random effects DL model in the metaviz R module. The overall effect size was d = 0.33, CI [0.09, 0.57].

Discussion

We conducted three pre-registered replications of the inaction inertia effect experiments by Tykocinski et al. (1995). Consistent with the findings in the target article, participants in the large-difference condition were less likely to act on the subsequent opportunity than participants in the small-difference condition. However, evidence regarding differences between the large-difference condition and the control condition was mixed. Furthermore, whereas the target article found no difference between the small-difference condition and the control condition, in our replication experiments, participants in the small-difference condition were more likely to act on the subsequent opportunity than participants in the control condition. The effects differed across the four experimental scenarios, with the strongest effects supporting inaction inertia found in the ski scenario and car scenario, followed by much weaker to null effects in the frequent flyer scenario, and a probably problematic control condition in the fitness scenario.

Robustness of findings

We conducted a series of robustness checks and the findings were largely consistent when we: (a) ran all analyses after applying pre-registered exclusion criteria, (b) included gender as a covariate, (c) used Welch's ANOVA rather than Fisher's ANOVA to analyze the data, and (d) used the restricted maximum-likelihood method to conduct meta-analysis. Details regarding these robustness checks are provided in supplementary materials.

Also, findings were largely consistent across the four samples. This is noteworthy because of the stark differences between the demographics of the Hong Kong undergraduate student sample and the two American MTurk online worker sample data collections (about 1 year apart). This also lends support to the robustness of the effects found, suggesting that both the successful and less successful effects are likely not due to random chance.

Replication findings evaluation

We concluded mostly successful replications in the ski and car scenarios. These scenarios can be categorized as having clear monetary implications, and relatively large differences between the numbers assigned to the differenet conditions (e.g., 500, 750, 2500).

We found weak to no effects in the frequent flyer scenario. Tykocinski et al. (1995) pointed out that "the frequent flyer scenario involved no monetary investment at all. Other than a trivial time investment associated with making a telephone call to the airline company, there were no costs contingent on joining the program" (pp. 796-797). Hence, one possibility for the weaker effects is that participants are less sensitive to non-monetary decisions, compared to the monetary decisions used in the ski and car scenarios. Another possibility is that the meaning of the materials used in the scenario has somehow changed over time. For example, in this case, it could be that 15,500 miles indicated in the scenario were valued highly in 1995, but that perceived value has depreciated over the years. Possibly, the numbers in this scenario could be adjusted and may result in stronger effects (e.g., 110,000 miles for a free international ticket, 85,000 miles for the large-difference condition, 40,000 miles for the small-difference condition, and 25,000 miles for the control condition). In failed replications there is often the urge to try and find some explanation for what might have gone wrong and suggest one of endless possibilities for potential remedies, yet regardless, the important takeaway is that researchers can no longer expect the scenario to work as expected as is.

In the fitness center scenario, we concluded support for the inaction inertia effect for the contrast between the large-difference and small-difference conditions, with findings overall as expected. However, the control condition showed an unexpected pattern of results, with a lower likelihood to act on a subsequent opportunity than participants in both the large and the small differences conditions. We suggest two possibilities. First, unlike the three other scenarios, the control condition in the fitness center scenario did not mention benefits (discounts or incentives) for joining a club, which may have reduced the attractiveness of the option. Without any reference information that allows comparison to the other conditions, the baseline rate depends on personal preference for fitness, and it could be that these were low in our samples. Future research may examine the perceived value of the control condition explicitly and adjust the parameters of the control condition to increase its attractiveness. For example, a more motivated control condition without a prior opportunity can be: “You have been considering becoming a member of a fitness club recently. Today you saw an enrolment advertisement of a fitness club located 30 min away. How likely are you to become a member of this fitness club now?”

Second, it is also possible that compared to a control condition with no prior opportunity, a small-difference condition offers a chance to compensate for a previously foregone, desirable opportunity. Thus, to avoid experiencing the regret of missing a similar or slightly less attractive opportunity for the second time, participants were more likely to act on the subsequent opportunity in the small-difference condition (Sevdalis, Harvey, & Yip, 2006; Shani, Danziger, & Zeelenberg, 2015).

Conclusion

We found support for the inaction inertia phenomenon, mostly consistent across four samples with unexpected nuances (mostly regarding contrasts against the control condition) and finding that contextual parameters seem to play a role. We recommend using the ski and car scenarios in future follow-up research on inaction inertia, and caution against using the frequent flyer and fitness scenarios as is without adjustments and further pre-testing. Future research should aim to better understand the differences between the different scenarios and the parameters affecting the strength of inaction inertia across situations.

References

Anderson, C. J. (2003). The psychology of doing nothing: forms of decision avoidance result from reason and emotion. *Psychological Bulletin*, *129*, 139-167.

Arkes, H. R., Kung, Y. H., & Hutzel, L. (2002). Regret, valuation, and inaction inertia. *Organizational Behavior and Human Decision Processes*, *87*, 371-385.

Bates, D., Maechler, M., Bolker, B. and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1-48.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., ... & Altmejd, A. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*, 637-644.

David Stanley (2018). apaTables: Create American Psychological Association (APA) Style Tables (R package version 2.0.5). Retrieved from https://CRAN.R-project.org/package=apaTables

DeCoster, J. (2012). *Converting effect sizes 2012-06-19.* Retrieved from www.stat-help.com/spreadsheets/Converting%20effect%20sizes%202012-06-19.xls

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Foster, C. C., & Diab, D. L. (2017). To accept or not to accept a job offer: Examining inaction inertia in an organizational context. *Journal of Applied Social Psychology*, *47*, 634-645.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. Unpublished manuscript. Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

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.

Kelley, K. (2007). Methods for the behavioral, educational, and social sciences: An R package. *Behavior Research Methods, 39,* 979-984.

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., ... & Batra, R. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, *1,* 443-490.

Kossmeier, M., Tran, U. S., & Voracek, M. (2019). *metaviz: Forest Plots, Funnel Plots, and Visual Funnel Plot Inference for Meta-Analysis* (R package version 0.3.0). Retrieved from https://CRAN.R-project.org/package=metaviz

Krijnen, J. M., Zeelenberg, M., Breugelmans, S. M., & van Putten, M. (2019). Inaction inertia in retirement saving. *Journal of Behavioral Decision Making*. Advance online publication. doi: 10.1002/bdm.2145

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4,* 863.

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.

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*, 389-402.

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.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., du Sert, N. P., ... & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*, 1-9.

Nosek, B. A., & Lakens, D. (2014). Registered reports: A Method to Increase the Credibility of Published Results. *Social Psychology, 45*, 137-141.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716.

Patil, I. (2018). ggstatsplot: 'ggplot2' based plots with statistical details. doi: 10.5281/zenodo.2074621. Retrieved from https://CRAN.R-project.org/package=ggstatsplot

Pittman, T. A., Tykocinski, O. E., Keinan, R.S., & Matthews, P.A. (2007). When bonuses backfire: an inaction inertia analysis of procrastination induced by a missed opportunity. *Journal of Behavioral Decision Making, 21*, 139-150.

R Core Team (2019). R: A language and environment for statistical computing [computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org

Sevdalis, N., Harvey, N., & Yip, M. (2006). Regret triggers inaction inertia–but which regret and how? *British Journal of Social Psychology, 45*, 839-853.

Shani, Y., Danziger, S., & Zeelenberg, M. (2015). Choosing between options associated with past and future regret. *Organizational Behavior and Human Decision Processes*, 126, 107–114.

Terris, L. G., & Tykocinski, O. E. (2016). Inaction inertia in international negotiations: The consequences of missed opportunities. *British Journal of Political Science*, *46*, 701-717.

Tykocinski, O. E., & Pittman, T. S. (1998). The consequences of doing nothing: Inaction inertia as avoidance of anticipated counterfactual regret. Journal of Personality and Social Psychology, 75, 607-616.

Tykocinski, O. E., & Pittman, T. S. (2001). Product aversion following a missed opportunity: price contrast or avoidance of anticipated regret? *Basic and Applied Social Psychology*, *23*, 149-156.

Tykocinski, O. E., Pittman, T. S., & Tuttle, E. E. (1995). Inaction inertia: Foregoing future benefits as a result of an initial failure to act. *Journal of Personality and Social Psychology*, *68*, 793-803.

Vosgerau, J., Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2019). 99% impossible: A valid, or falsifiable, internal meta-analysis. *Journal of Experimental Psychology: General*, *148*, 1628 -1639.

Inaction Inertia replications: Supplementary Materials

Disclosures

Procedure

The replication conducted with Hong Kong students was conducted as part of a large replication project, where students participated voluntarily in a survey where we attempted to replicate several findings from the judgment and decision-making literature.

Similarly, the replication conducted with Amazon Mechanical Turk workers was conducted as part of a large replication project, where AMT workers were paid a nominal amount to participate in a survey where we attempted to replicate several findings from the judgment and decision-making literature.

Pre-registrations

Pre-registrations were conducted prior to data collection.

Links:

1. Experiment 1 pre-registration: https://osf.io/kbnjw;
2. Experiment 2 pre-registration: https://osf.io/e93k4;
3. Experiment 3 pre-registration: https://osf.io/83w2x

Data collection

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

Data collections for Experiments 1 and 2 were combined with other independent replication attempts unrelated to this replication and displayed in randomized order.

Conditions reporting

All collected conditions are reported.

Data exclusions

In the Hong Kong sample, the data collection was performed on student designing replications. We excluded participants who designed the study.

We note that the pre-registration plans included different references to possible exclusion criteria addressing seriousness, English proficiency, etc. We conducted our analyses both with and without exclusions, and found that exclusions had little effect on the results, the undergraduate samples and MTurk samples were proficient and serious. For the sake of brevity, the findings reported in the manuscript are without any further exclusions.

Variable reporting

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

Open science

Datasets, code, and supplementary materials were made available on the Open Science Framework at: https://osf.io/kxe73/ .

Materials

Ski scenario

Large differences

Imagine the following scenario: Your friend called you at the beginning of October and told you that he intended to buy a special pass to Ski Liberty (where you both like to ski). He said that the deal was that if you bought the pass before the 15th of October the pass would cost you only **$40** instead of the **$100** regular price. Although it sounded like a good idea, you forgot to do it by the 15th. The next time your friend called he told you that although you missed the deadline you could still get a pass for **$90** if you pay this week.
How likely are you to spend the **$90**?

0 - *Not at all likely* to 10 - *Extremely likely*

Small differences

Imagine the following scenario: Your friend called you at the beginning of October and told you that he intended to buy a special pass to Ski Liberty (where you both like to ski). He said that the deal was that if you bought the pass before the 15th of October the pass would cost you only **$80** instead of the **$100** regular price. Although it sounded like a good idea, you forgot to do it by the 15th. The next time your friend called he told you that although you missed the deadline you could still get a pass for **$90** if you pay this week.
How likely are you to spend the **$90**?

0 - *Not at all likely* to 10 - *Extremely likely*

Control condition (no foregone alternative)

Imagine the following scenario: Your friend called you at the beginning of October and told you that he intended to buy a special pass to Ski Liberty (where you both like to ski). He said that the deal was that if you buy the pass this week, it will only cost you **$90** instead of the **$100** regular price.
How likely are you to spend the **$90**?

0 - *Not at all likely* to 10 - *Extremely likely*

Car scenario

Large differences

You see a television advertisement by a local dealer for a car and you are interested in buying. The advertisement promotes a limited-time **$500** factory rebate on the car, providing it is purchased this week. However, you had seen this advertisement once before. Back then, the car was offered with a larger rebate value, **$2,500**, for a limited time. Although you were interested in the deal at that time, you had missed the deadline.
How likely are you to buy the car this week?

0 - *Not at all likely* to 10 - *Extremely likely*

Small differences

Imagine the following scenario: You see a television advertisement by a local dealer for a car and you are interested in buying. The advertisement promotes a limited-time **$500** factory rebate on the car, providing it is purchased this week. However, you had seen this advertisement once before. Back then, the car was

offered with a larger rebate value, **$750**, for a limited time. Although you were interested in the deal at that time, you had missed the deadline.
How likely are you to buy the car this week?

0 - *Not at all likely* to 10 - *Extremely likely*

Control condition (no foregone alternative)

Imagine the following scenario: You see a television advertisement by a local dealer for a car and you are interested in buying. The advertisement promotes a limited-time **$500** factory rebate on the car, providing it is purchased this week.
How likely are you to buy the car this week?

0 - *Not at all likely* to 10 - *Extremely likely*

Flyer scenario

Large differences

Imagine the following scenario: You are considering joining a frequent flyer program before taking a trip for the holidays. If you decide to join, you would accumulate **5,500** miles (towards a free ticket, on accumulating **20,000** miles). You had considered joining this program once before, near the beginning of the year, but had not done so. The number of miles that would have been accumulated after this current trip had you joined earlier was **15,500** miles.
How likely are you to join the program now?

0 - *Not at all likely* to 10 - *Extremely likely*

Small differences

Imagine the following scenario: You are considering joining a frequent flyer program before taking a trip for the holidays. If you decide to join, you would accumulate **5,500** miles (towards a free ticket, on accumulating **20,000** miles). You had considered joining this program once before, near the beginning of the year, but had not done so. The number of miles that would have been accumulated after this current trip had you joined earlier was **7,500** miles.
How likely are you to join the program now?

0 - *Not at all likely* to 10 - *Extremely likely*

Control condition (no foregone alternative)

Imagine the following scenario: You are considering joining a frequent flyer program before taking a trip for the holidays. If you decide to join, you would accumulate **5,500** miles (towards a free ticket, on accumulating **20,000** miles).
How likely are you to join the program?

0 - *Not at all likely* to 10 - *Extremely likely*

Fitness scenario

Large differences

Imagine the following scenario: You are considering becoming a member of a fitness club located **30 min away**. You could have joined another center closer to home, but, having neglected to act quickly, had missed the opportunity. The center that by now had closed its membership roll was said to be located **5**

<u>min away</u> from your home.
How likely are you to become a member of the fitness club now?

0 - *Not at all likely* to 10 - *Extremely likely*

Small differences

Imagine the following scenario: You are considering becoming a member of a fitness club located **<u>30 min away</u>**. You could have joined another center closer to home, but, having neglected to act quickly, had missed the opportunity. The center that by now had closed its membership roll was said to be located **<u>25 min away</u>** from your home.
How likely are you to become a member of the fitness club now?

0 - *Not at all likely* to 10 - *Extremely likely*

Control condition (no foregone alternative)

Imagine the following scenario: You are considering becoming a member of a fitness club located **<u>30 min away</u>**.
How likely are you to become a member of the fitness club now?

0 - *Not at all likely* to 10 - *Extremely likely*

Project Process Outline

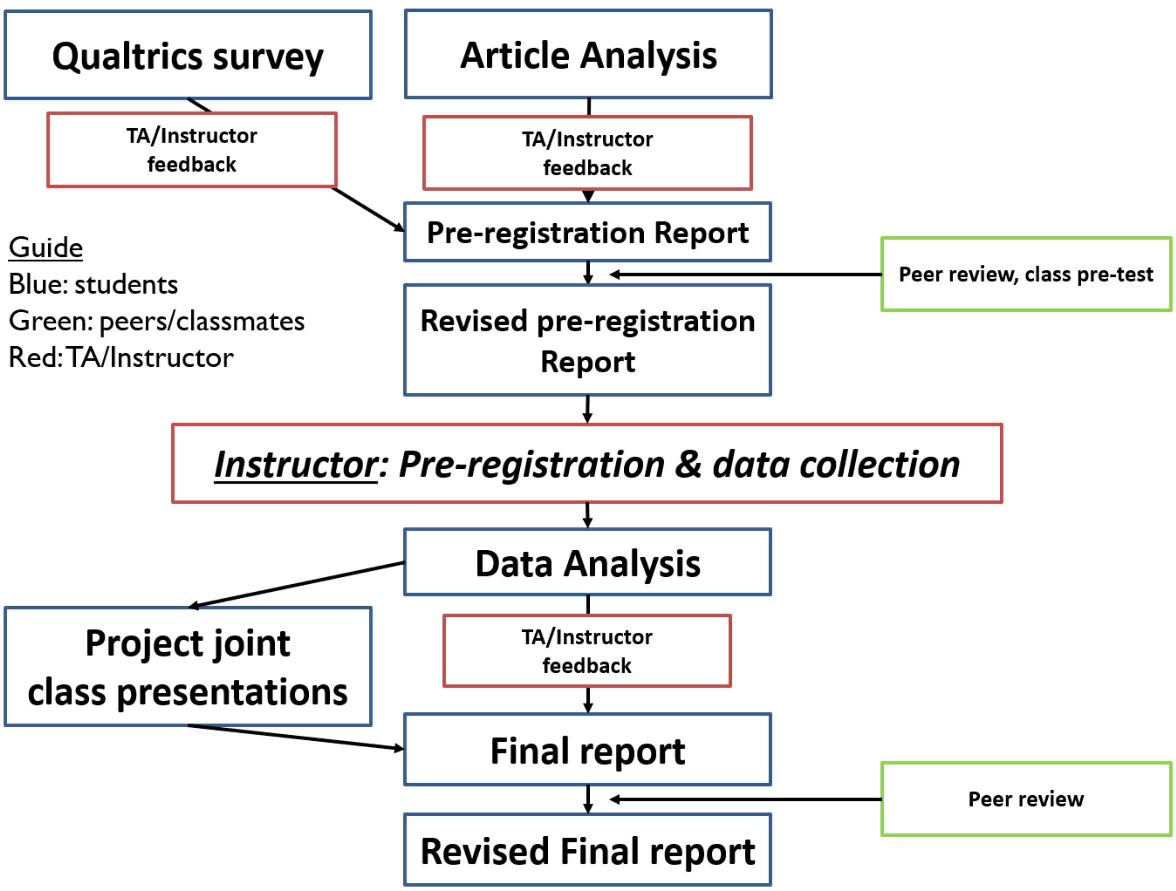
Verification of Analyses

Initial analyses were conducted by student replicators, who were to use jamovi (jamovi project, 2018) in the analyses.

In preparing this manuscript, the lead author and the corresponding author verified analyses in SPSS (IBM Corp, 2017) and R ggstatsplot, and the results across platforms were consistent. This is the main reason for the mixed code in the OSF.

Confidence intervals for eta-squared were calculated using the command get.ci.partial.eta.squared in the R package apaTables (Stanley, 2018). Point estimates and confidence intervals for Cohen's *d* were calculated using the command ci.smd in the R package MBESS (Kelley, 2018).

Advanced social psychology course flow-diagram



Effect sizes calculations and Power Analyses

<u>Ski Scenario</u>

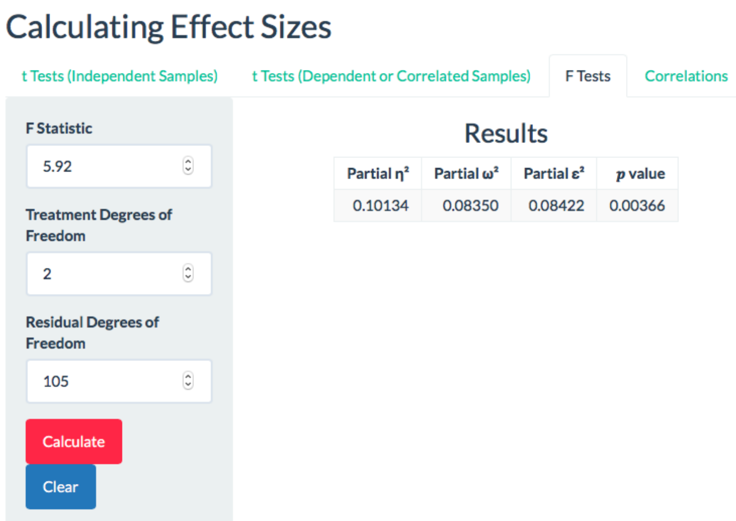
One-way ANOVA of the Target Article:

n = 108, $F(2, 105) = 5.92$, $p < .004$

Calculation of the eta-squared using F-statistics

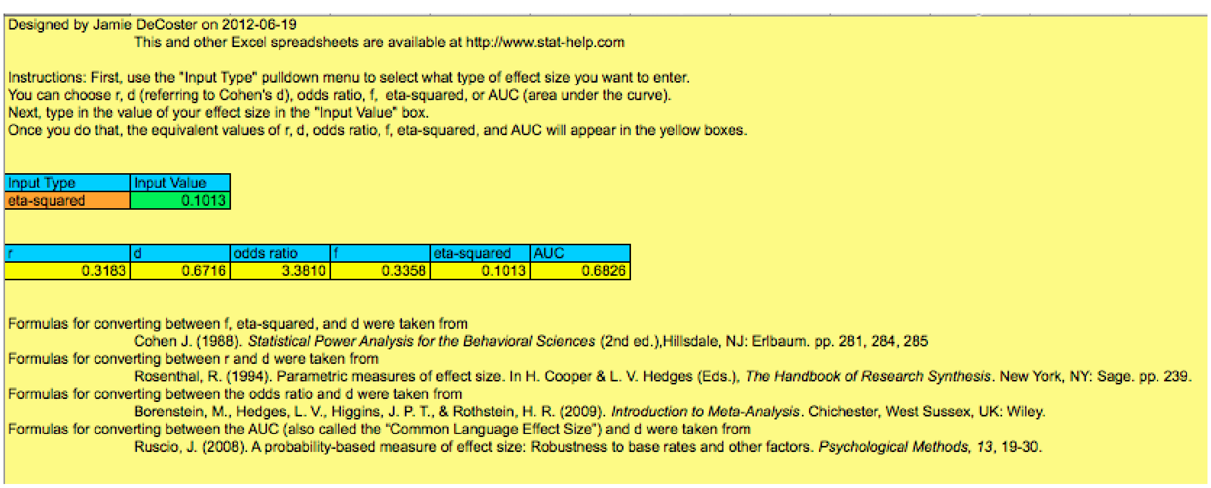
Using: https://katherinemwood.shinyapps.io/lakens_effect_sizes/

Partial $\eta^2 = 0.10134$



Calculation of the effect size using eta-squared

Effect size $f = 0.3358$



Designed by Jamie DeCoster on 2012-06-19
This and other Excel spreadsheets are available at http://www.stat-help.com

Instructions: First, use the "Input Type" pulldown menu to select what type of effect size you want to enter.
You can choose r, d (referring to Cohen's d), odds ratio, f, eta-squared, or AUC (area under the curve).
Next, type in the value of your effect size in the "Input Value" box.
Once you do that, the equivalent values of r, d, odds ratio, f, eta-squared, and AUC will appear in the yellow boxes.

Input Type	Input Value
eta-squared	0.1013

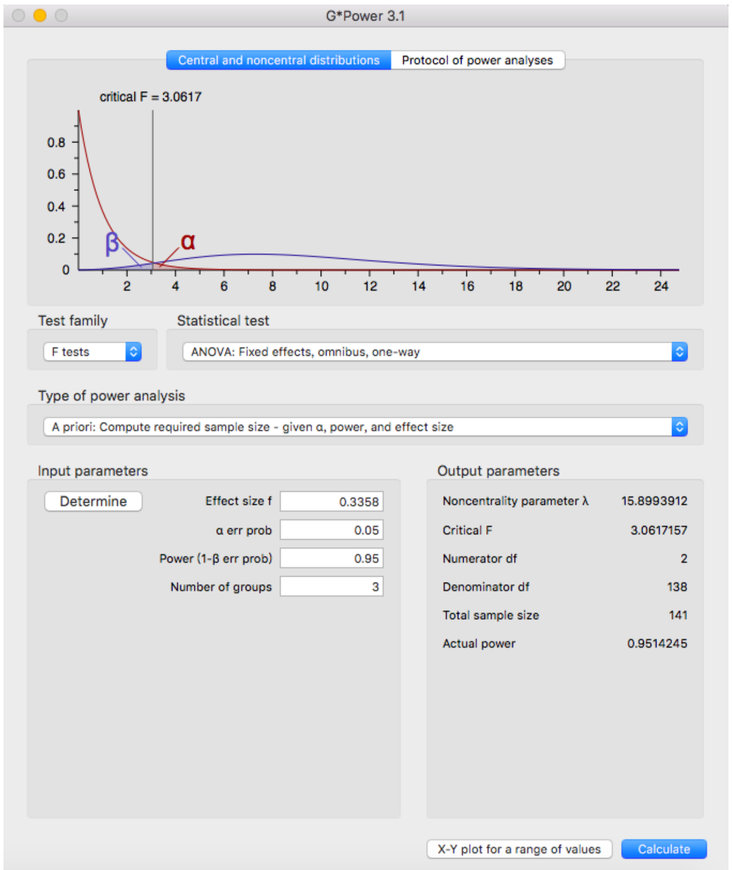
r	d	odds ratio	f	eta-squared	AUC
0.3183	0.6716	3.3810	0.3358	0.1013	0.6826

Formulas for converting between f, eta-squared, and d were taken from
Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.),Hillsdale, NJ: Erlbaum. pp. 281, 284, 285
Formulas for converting between r and d were taken from
Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis*. New York, NY: Sage. pp. 239.
Formulas for converting between the odds ratio and d were taken from
Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, West Sussex, UK: Wiley.
Formulas for converting between the AUC (also called the "Common Language Effect Size") and d were taken from
Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods, 13*, 19-30.

Calculation of the required sample size using effect size

Using: G*Power Version 3.1.9.2

Required *n* = 141



F tests - ANOVA: Fixed effects, omnibus, one-way

Analysis:	A priori: Compute required sample size		
Input:	Effect size f	=	0.3358
	α err prob	=	0.05
	Power (1-β err prob)	=	0.95
	Number of groups	=	3
Output:	Noncentrality parameter λ	=	15.8993912
	Critical F	=	3.0617157
	Numerator df	=	2
	Denominator df	=	138
	Total sample size	=	141
	Actual power	=	0.9514245

Car Scenario

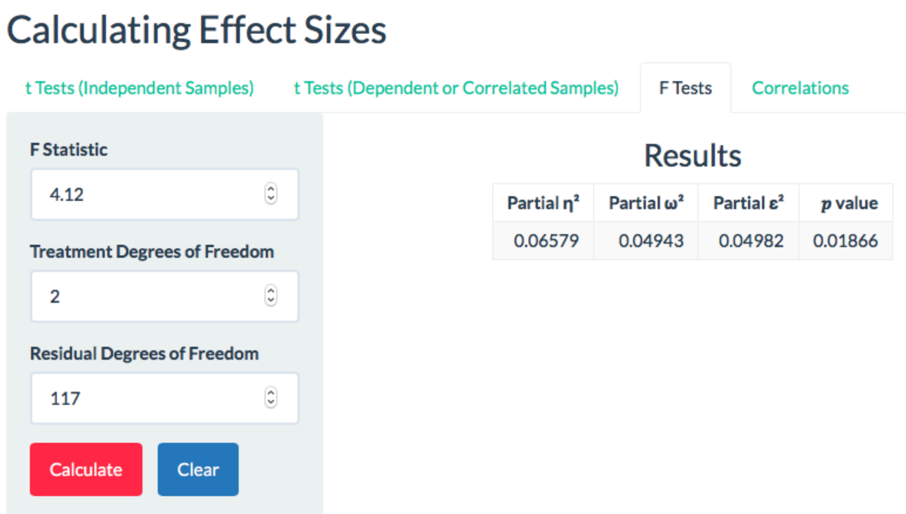
One-way ANOVA of the target article

n = 120, $F(2, 117) = 4.12$, $p < .02$

Calculation of the eta-squared using F-statistics

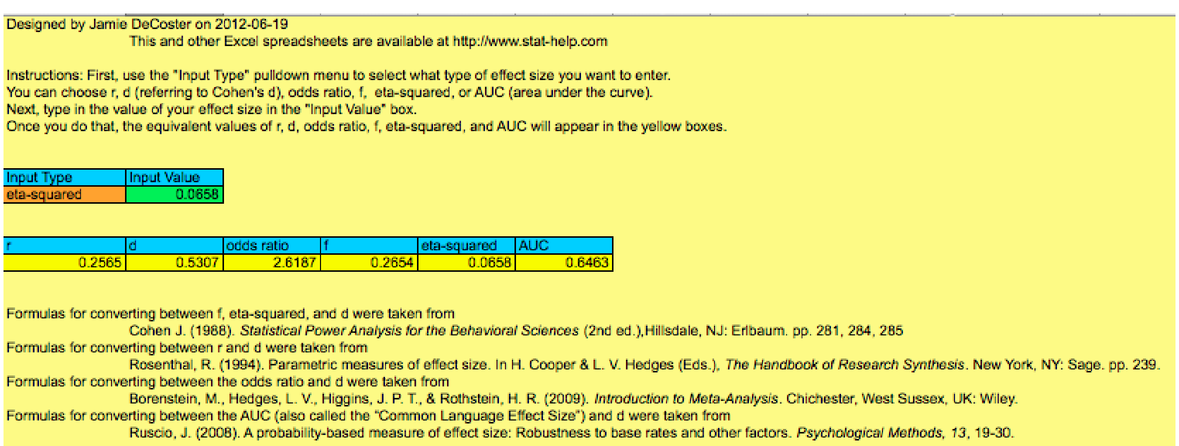
Using: https://katherinemwood.shinyapps.io/lakens_effect_sizes/

Partial $\eta^2 = 0.06579$



Calculation of the effect size using eta-squared

Effect size f = 0.2654



Designed by Jamie DeCoster on 2012-06-19
This and other Excel spreadsheets are available at http://www.stat-help.com

Instructions: First, use the "Input Type" pulldown menu to select what type of effect size you want to enter.
You can choose r, d (referring to Cohen's d), odds ratio, f, eta-squared, or AUC (area under the curve).
Next, type in the value of your effect size in the "Input Value" box.
Once you do that, the equivalent values of r, d, odds ratio, f, eta-squared, and AUC will appear in the yellow boxes.

Input Type	Input Value
eta-squared	0.0658

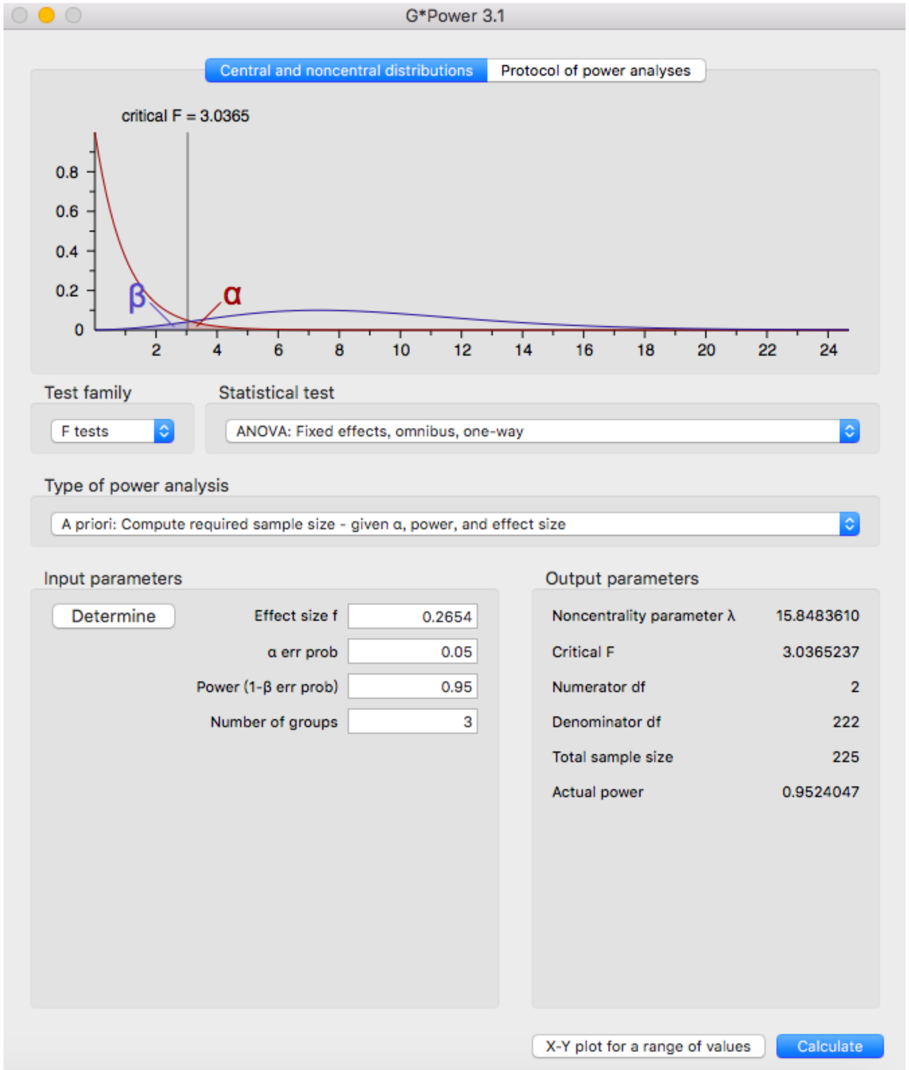
r	d	odds ratio	f	eta-squared	AUC
0.2565	0.5307	2.6187	0.2654	0.0658	0.6463

Formulas for converting between f, eta-squared, and d were taken from
Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.),Hillsdale, NJ: Erlbaum. pp. 281, 284, 285
Formulas for converting between r and d were taken from
Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis*. New York, NY: Sage. pp. 239.
Formulas for converting between the odds ratio and d were taken from
Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, West Sussex, UK: Wiley.
Formulas for converting between the AUC (also called the "Common Language Effect Size") and d were taken from
Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods*, *13*, 19-30.

Calculation of the required sample size using effect size

Using: G*Power Version 3.1.9.2

Required n = 225




```
F tests - ANOVA: Fixed effects, omnibus, one-way

Analysis:   A priori: Compute required sample size
Input:      Effect size f                  =   0.2654
            α err prob                     =   0.05
            Power (1-β err prob)           =   0.95
            Number of groups               =   3
Output:     Noncentrality parameter λ      =   15.8483610
            Critical F                     =   3.0365237
            Numerator df                   =   2
            Denominator df                 =   222
            Total sample size              =   225
            Actual power                   =   0.9524047
```

<u>Frequent Flyer Scenario</u>

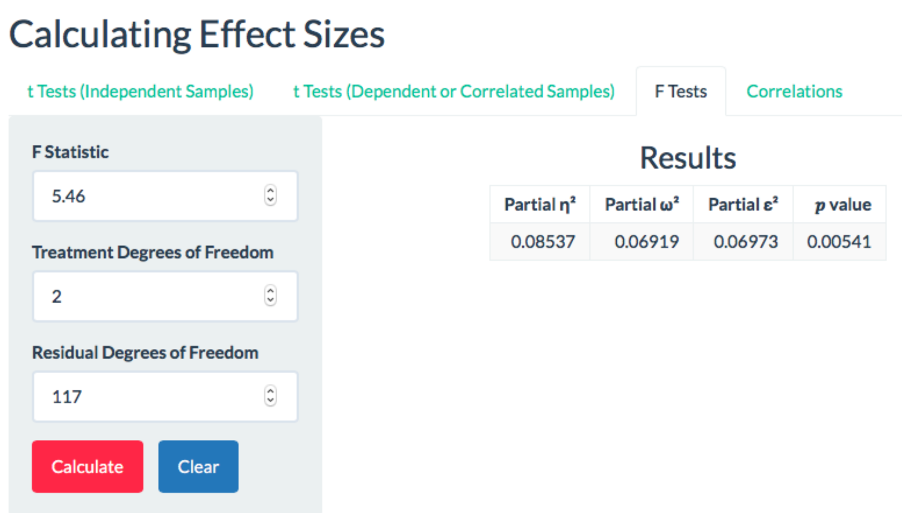
One-way ANOVA of the target article

n = 120, $F(2, 117) = 5.46$, $p < .006$

Calculation of the eta-squared using F-statistics

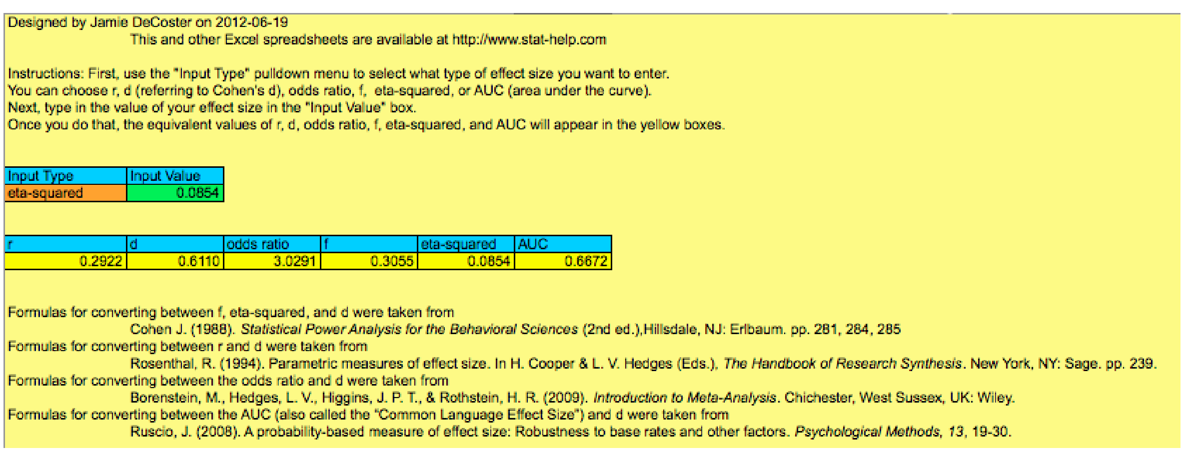
Using: https://katherinemwood.shinyapps.io/lakens_effect_sizes/

Partial $\eta^2 = 0.08537$



Calculation of the effect size using eta-squared

Effect size f = 0.3055



Designed by Jamie DeCoster on 2012-06-19
This and other Excel spreadsheets are available at http://www.stat-help.com

Instructions: First, use the "Input Type" pulldown menu to select what type of effect size you want to enter.
You can choose r, d (referring to Cohen's d), odds ratio, f, eta-squared, or AUC (area under the curve).
Next, type in the value of your effect size in the "Input Value" box.
Once you do that, the equivalent values of r, d, odds ratio, f, eta-squared, and AUC will appear in the yellow boxes.

Input Type	Input Value
eta-squared	0.0854

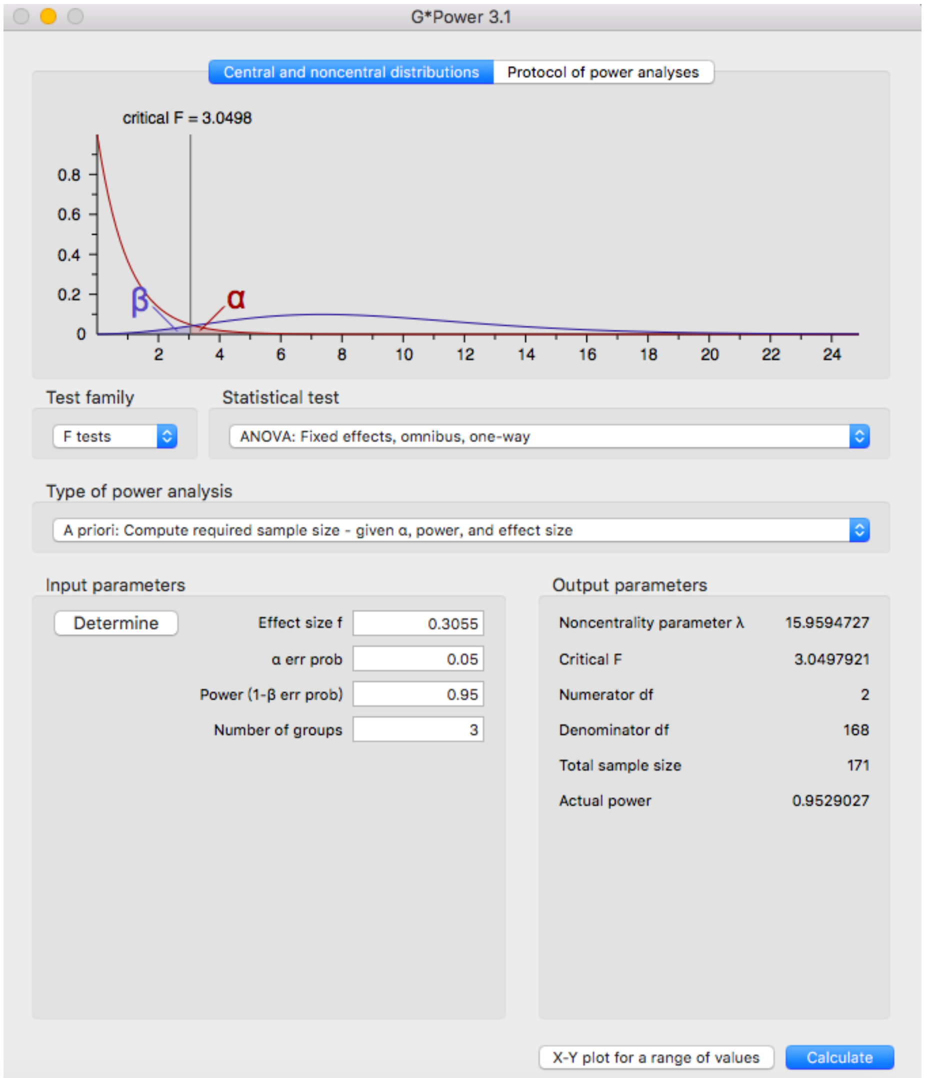
r	d	odds ratio	f	eta-squared	AUC
0.2922	0.6110	3.0291	0.3055	0.0854	0.6672

Formulas for converting between f, eta-squared, and d were taken from
Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.),Hillsdale, NJ: Erlbaum. pp. 281, 284, 285
Formulas for converting between r and d were taken from
Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis*. New York, NY: Sage. pp. 239.
Formulas for converting between the odds ratio and d were taken from
Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, West Sussex, UK: Wiley.
Formulas for converting between the AUC (also called the "Common Language Effect Size") and d were taken from
Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods*, *13*, 19-30.

Calculation of the required sample size using effect size

Using: G*Power Version 3.1.9.2

Required *n* = 171




```
F tests - ANOVA: Fixed effects, omnibus, one-way

Analysis:   A priori: Compute required sample size
Input:      Effect size f                  =   0.3055
            α err prob                     =   0.05
            Power (1-β err prob)           =   0.95
            Number of groups               =   3
Output:     Noncentrality parameter λ      =   15.9594727
            Critical F                     =   3.0497921
            Numerator df                   =   2
            Denominator df                 =   168
            Total sample size              =   171
            Actual power                   =   0.9529027
```

Fitness Center Scenario

One-way ANOVA of the target article

n = 120, $F(2, 117) = 3.92$, $p < .03$

Calculation of the minium sample size

Because the fitness center scenario has the smallest effect size among the four scenarios, it serves as the basis for calculating the minimum required sample size.

Power: $1\text{-}\beta = 0.95$

Significance: $\alpha = 0.05$

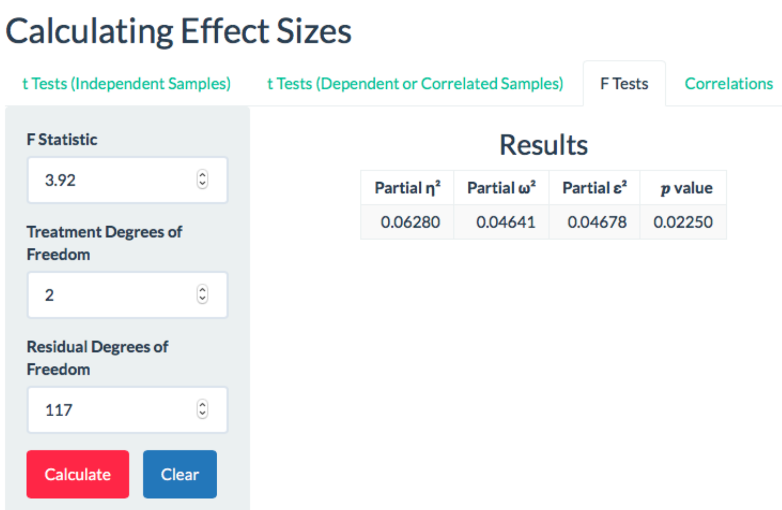
Effect size: $f = 0.2589$

Required n = 234

Calculation of the eta-squared using F-statistics

Using: https://katherinemwood.shinyapps.io/lakens_effect_sizes/

Partial η^2 = .06280



Calculation of the effect size using eta-squared

Effect size f = 0.2589

Designed by Jamie DeCoster on 2012-06-19
This and other Excel spreadsheets are available at http://www.stat-help.com

Instructions: First, use the "Input Type" pulldown menu to select what type of effect size you want to enter.
You can choose r, d (referring to Cohen's d), odds ratio, f, eta-squared, or AUC (area under the curve).
Next, type in the value of your effect size in the "Input Value" box.
Once you do that, the equivalent values of r, d, odds ratio, f, eta-squared, and AUC will appear in the yellow boxes.

Input Type	Input Value
eta-squared	0.0628

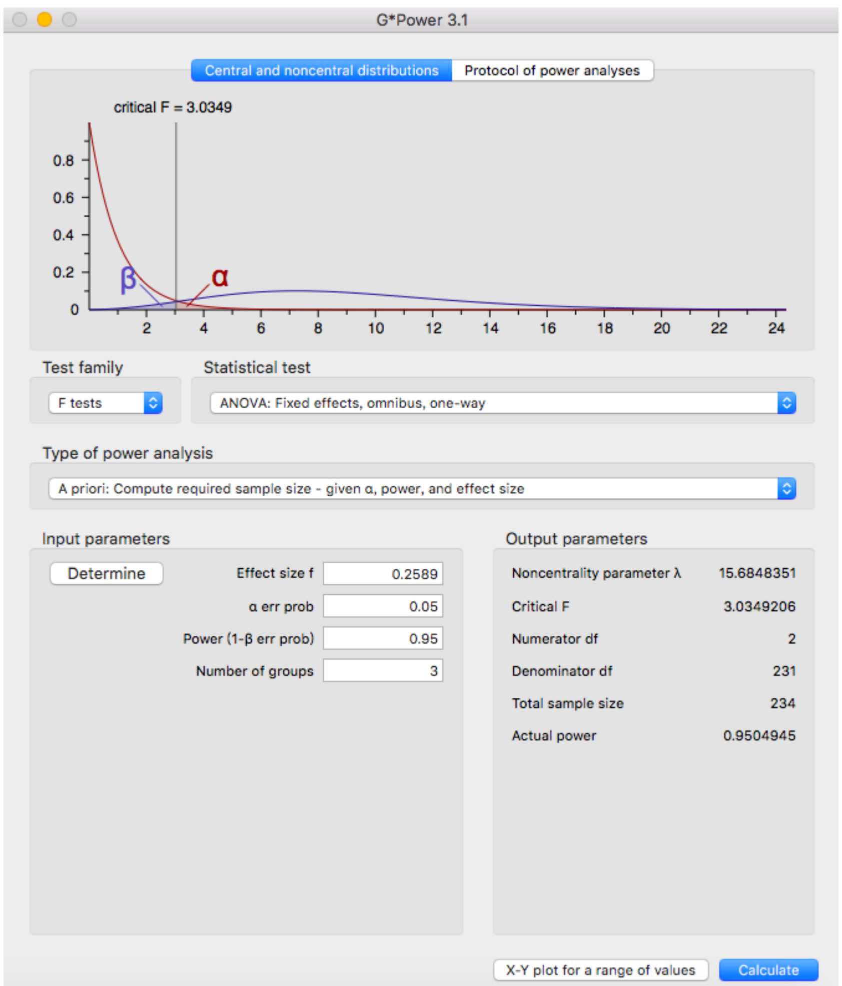
r	d	odds ratio	f	eta-squared	AUC
0.2506	0.5177	2.5575	0.2589	0.0628	0.6428

Formulas for converting between f, eta-squared, and d were taken from
Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.),Hillsdale, NJ: Erlbaum. pp. 281, 284, 285
Formulas for converting between r and d were taken from
Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis*. New York, NY: Sage. pp. 239.
Formulas for converting between the odds ratio and d were taken from
Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, West Sussex, UK: Wiley.
Formulas for converting between the AUC (also called the "Common Language Effect Size") and d were taken from
Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods, 13*, 19-30.

Calculation of the required sample size using effect size

Using: G*Power Version 3.1.9.2

Required n = 234



F tests - ANOVA: Fixed effects, omnibus, one-way

Analysis:	A priori: Compute required sample size		
Input:	Effect size f	=	0.2589
	α err prob	=	0.05
	Power (1-β err prob)	=	0.95
	Number of groups	=	3
Output:	Noncentrality parameter λ	=	15.6848351
	Critical F	=	3.0349206
	Numerator df	=	2
	Denominator df	=	231
	Total sample size	=	234
	Actual power	=	0.9504945

Classification of the Three Replication Studies Based on Lebel et al.'s (2018) Taxonomy

We summarize an evaluation of the three replications in terms of similarly to the experiments in the target article using the taxonomy introduced by LeBel, McCarthy, Earp, Elson, and Vanpaemel (2018). In this taxonomy, replications range from exact replication to very far replication. Key design facets for evaluating methodological similarity include operationalizations of independent variables and dependent variables, stimuli used to measure independent variables and dependent variables, procedural details, physical settings, and contextual variables.

Adopting this taxonomy, the three experiments adhered to the criteria of very close replications. All design aspects were the same as in the target article except for procedural details and physical setting. See Table A1 for more details.

Table A1

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

Design facet	**Experiment 1** **Hong Kong replication**	**Experient2** **MTurk replication**	**Experient3** **MTurk replication**
Effect, hypotheses	same	same	same
IV construct	same	same	same
DV construct	same	same	same
IV operationalization	same	same	same
DV operationalization	same	same	same
IV stimuli	same	same	same
DV stimuli	same	same	same
Procedural details	different	different	different
Physical settings	different	different	different
Contextual variables	different	different	different
Replication classification	Very close replication	Very close replication	Very close replication

Sample Demographic Information

Experiment 1 - Hong Kong undergraduate sample

The first replication was considered a pre-test in an undergraduate course at a university in Hong Kong. Students were randomly assigned to work in group of 3 to 6 people, with each group responsible for conducting one replication, and one of the groups was in charge of the current replication. The students then served as the target sample for the experiments designed by their classmates, which they had no knowledge of prior to participation. The course materials covered classic judgement and decision-making literature, which meant that students were made aware of a wide array of heuristics and biases. The experiment therefore can be considered as a very conservative test of the effect in a non-naive sample.

The student designed the experiment survey, conducted effect size calculations and power analyses, and wrote the pre-registration. Pre-registration on the OSF and data collection were managed by the course instructor, the corresponding author. All students registered in the course were invited to take part as respondents in the study. To ensure anonymity, students were only asked to indicate which replication group they belonged to and those were later excluded from the data analysis of the study they designed. The final sample included 43 student participants (13 men, 30 women; M_{age} = 20.20, SD_{age} = 1.00). The majority of the participants were originally from Hong Kong SAR (34 students, 79.07%), two (4.65%) were from mainland China, and seven (16.28%) were international students.

Experiment 2 - online American Amazon Mechanical Turk

A total of 309 participants were recruited online from MTurk using Turkprime.com platform (Litman, Robinson, & Abberbock, 2017) (140 men, 169 women; M_{age} = 38.42, SD_{age} = 11.53). The majority of the participants were originally from the United States (305 people, 98.71%), and the other participants came from India, Philippines, and Ukraine, or did not report their country of origin. In terms of social class, 7.44% self-identified as from a lower class, 24.92% working class, 16.83% lower middle class, 40.45% middle class, 9.39% upper middle class, and 0.98% upper class.

Experiment 3 - online American Amazon Mechanical Turk

A total of 1203 participants were recruited online from MTurk using Turkprime.com platform (Litman et al., 2017). We randomly assigned participants into one of the two designs: different conditions design (603 participants) and same condition design (600 participants). In the different conditions design, participants were randomly assigned to one of the conditions (i.e., large-difference, small-difference, control) every time they finished reading a scenario, thus they could be assigned to different conditions in different scenarios. This is also the design used in Experiments 1 and 2. In the same condition design, a participant will be consistently assigned to the same condition in all scenarios, which is consistent with a typical mixed factorial design.

In different conditions sample, 263 of the participants are men, and 340 are women. Their mean age was 40.40 years old (SD_{age} = 12.25). We asked for an indication of country of birth. The majority of the participants were born in the United States (590 people, 97.84%), and the other participants were born in Bulgaria, Cabo Verde, Cambodia, Canada, Germany, Philippines, and United Kingdom, or did not report their country of origin. In terms of social class, 5.31% self-identified as from a lower class, 20.07% working class, 18.74% lower middle class, 45.61% middle class, 9.62% upper middle class, and 0.66% upper class.

In the same condition sample, 299 were men, and 301 were women. Their mean age was 40.40 years old (SD_{age} = 12.19). The majority of the participants were originally from the United States (590 people, 97.84%), and the other participants came from Germany and South Korea, or did not report their country of origin. In terms of social class, 3.00% self-identified as from a lower class, 23.00% working class, 19.33% lower middle class, 42.00% middle class, 12.17% upper middle class, and 0.50 % upper class.

Statistical assumptions and normality Tests (Q-Q Plots)

We conducted a series of tests of statistical assumptions for analyses. These tests include: a) homogeneity of variance (using Levene's test), and b) normality of residuals (using Q-Q plot).

Experiment 1

We ran the Levene's test for all four scenarios. The assumption of homogeneity of variance held for Scenario 1 (F (2, 40) = .46, p = .64), Scenario 2 (F (2, 40) = .28, p = .76), and Scenario 3 (F (2, 40) = 2.33, p = .11), but not for Scenario 4 (F (2, 40) = 5.14, p = .01). As indicated in our pre-registration, we would use the Games-Howell test rather than Tukey's test or Hochberg's GT2 test for pairwise comparisons of Scenario 4. The Games-Howell test is appropriate when variances across groups are unequal and the sample sizes per condition are unequal.

In order to test the normality of residuals, we ran Q-Q plot analyses for all four scenarios. The assumption of normality of residuals were satisfied in most of the conditions, but not satisfied in the following conditions: Scenario 2 large-difference condition (F (15) = .22, p = .04), Scenario 3 control condition (F (15) = .24, p = .02), and Scenario 4 small-difference condition (F (15) = .25, p = .02).

Experiment 2

We ran the Levene's test for all four scenarios. The assumption of homogeneity of variance held for Scenario 1 (F (2, 306) = .17, p = .85), Scenario 3 (F (2, 306) = 1.49, p = .23), and Scenario 4 (F (2, 306) = .04, p = .96), but not for Scenario 2 (F (2, 306) = 7.66, p = .00). Therefore, as indicated in our pre-registration, we would use the Games-Howell test rather than Tukey's test for the pairwise comparisons of Scenario 2.

In order to test the normality of residuals, we ran Q-Q plot analyses for all four scenarios. In summary, the assumption of normality of residuals was not satisfied in any condition in any scenario.

Experiment 3 (Different Conditions Design)

We ran the Levene's test for all four scenarios. The assumption of homogeneity of variance held for Scenario 3 (F (2, 600) = .27, p = .77), but not for Scenario 1 (F (2, 600) = 3.97, p = .02), Scenario 2 (F (2, 600) = .3.41, p = .03), and Scenario 4 (F (2, 600) = 13.60, p < .001). Therefore, as indicated in our pre-registration, we would use the Games-Howell test rather than Tukey's test for the pairwise comparisons of all scenarios except Scenario 3.

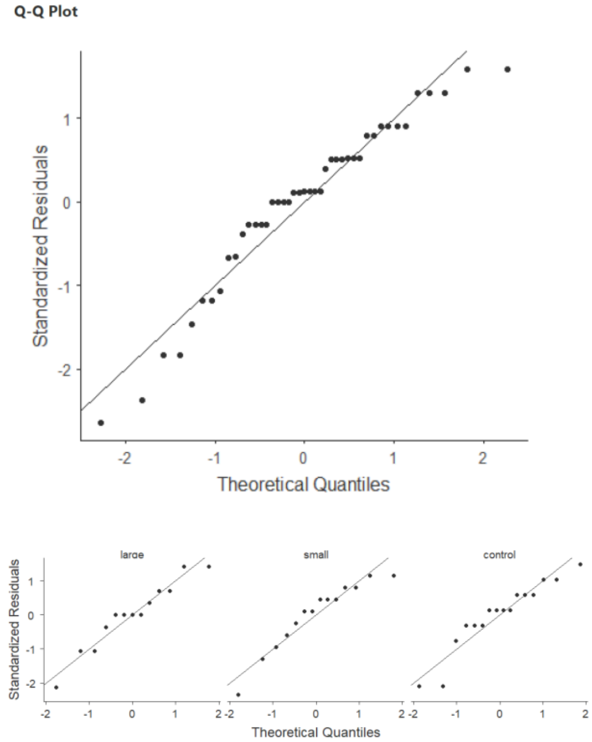
In order to test the normality of residuals, we ran Q-Q plot analyses for all four scenarios. In summary, the assumption of normality of residuals was not satisfied in any condition in any scenario.

Experiment 3 (Same Condition Design)

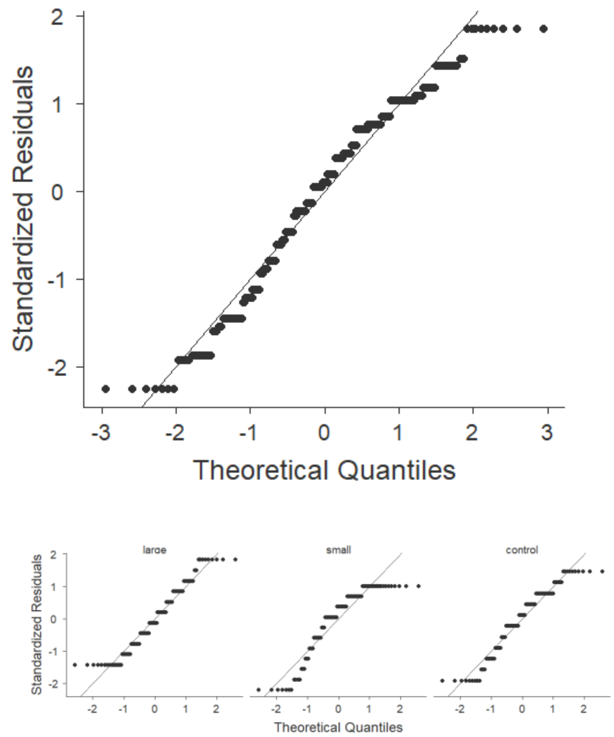
We ran the Levene's test for all four scenarios. The assumption of homogeneity of variance did not hold for any of the scenarios: Scenario 1 (F (2, 597) = 10.50, p < .001), Scenario 2 (F (2, 597) = 3.69, p = .03), Scenario 3 (F (2, 597) = 4.71, p = .01), Scenario 4 (F (2, 597) = 12.49, p < .001). Therefore, as indicated in our pre-registration, we would use the Games-Howell test rather than Tukey's test for the pairwise comparisons of all scenarios.

In order to test the normality of residuals, we ran Q-Q plot analyses for all four scenarios. In summary, the assumption of normality of residuals was not satisfied in any condition in any scenario.

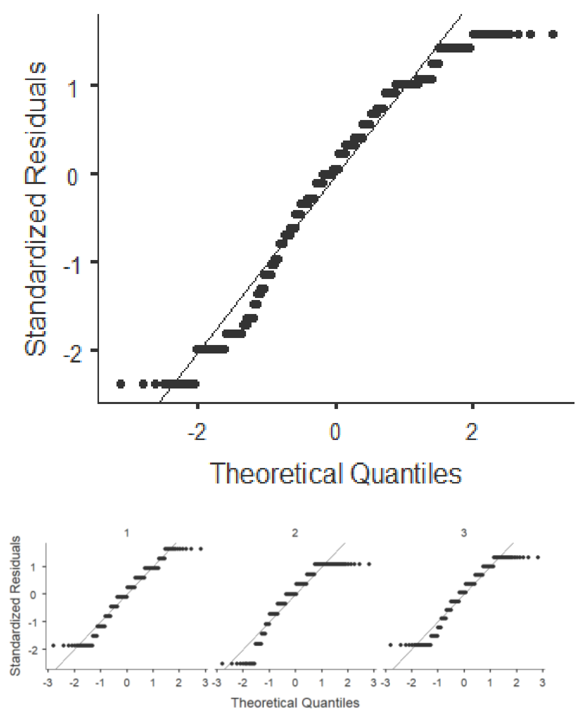
Ski Scenario



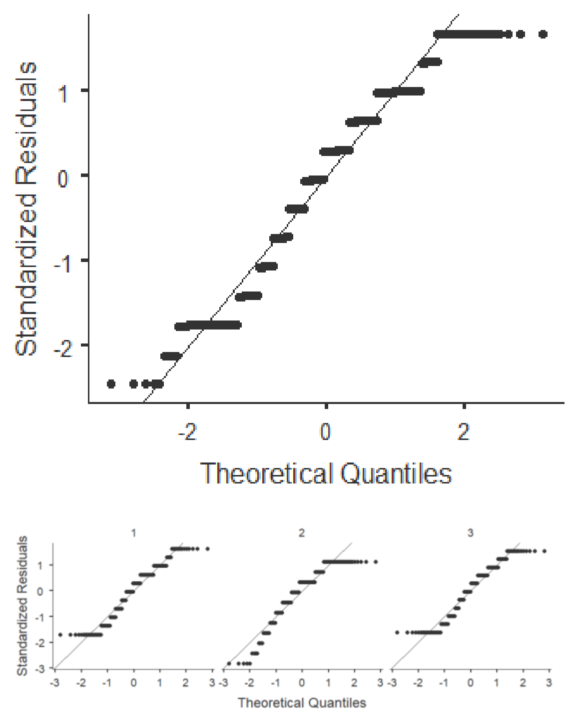
Experiment 1



Experiment 2



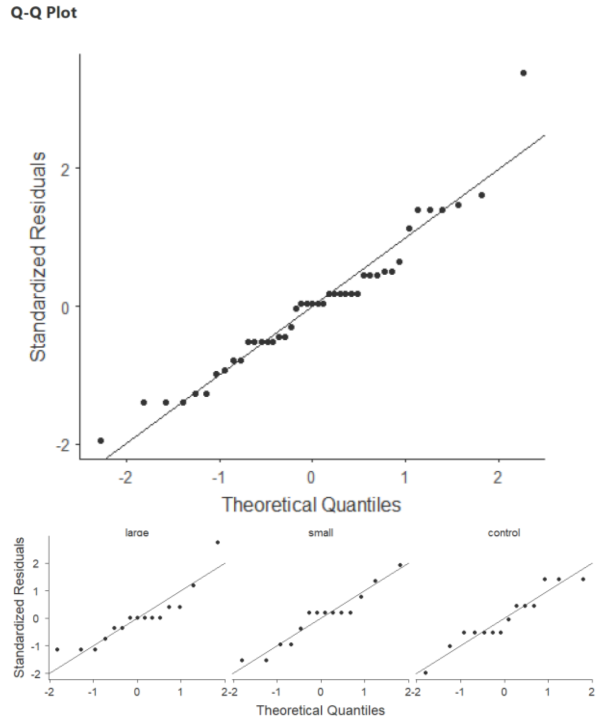
Experiment 3 (Different Conditions)



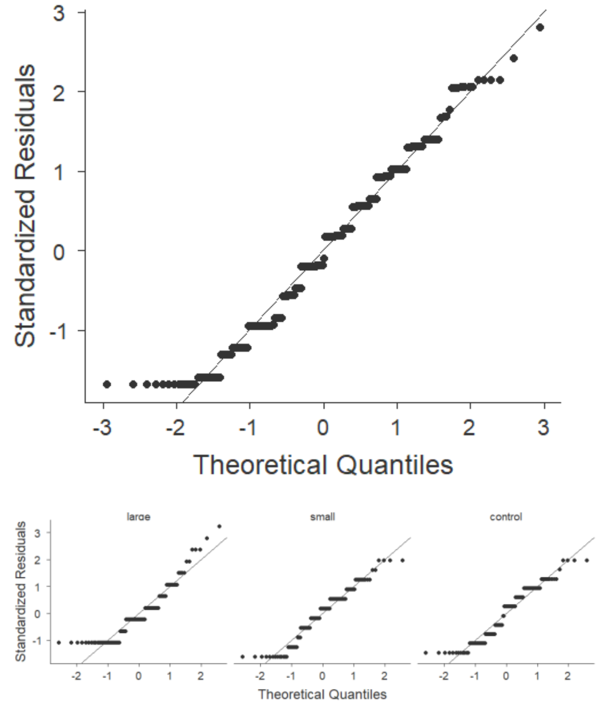
Experiment 4 (Same Condition)

Figure C1. Q-Q Plots of the four samples overall and in large, small, and control conditions (ski scenario).

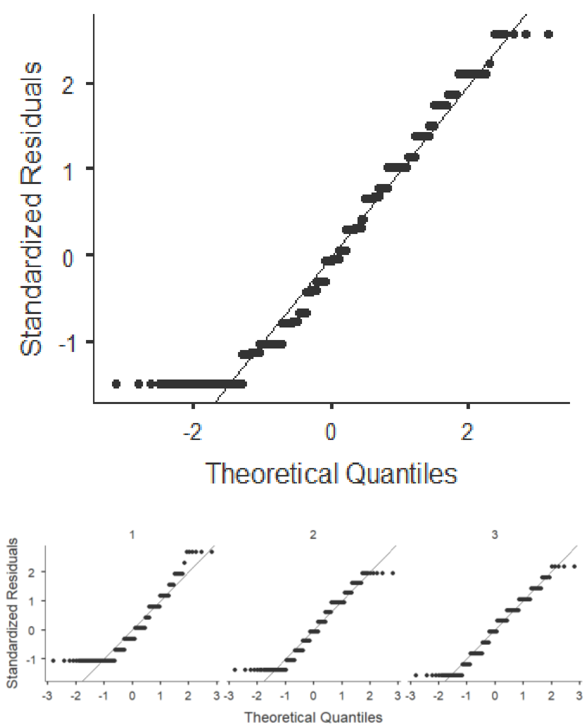
<u>Car Scenario</u>



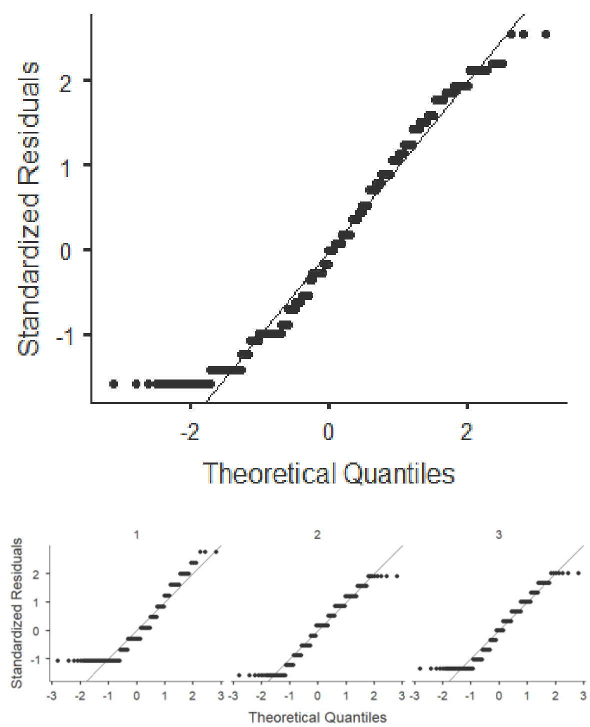
Experiment 1



Experiment 2



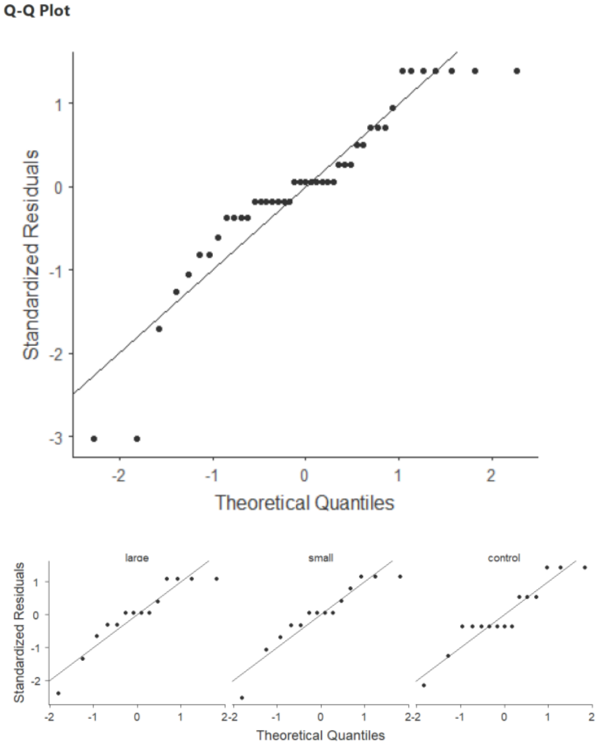
Experiment 3 (Different Conditions)



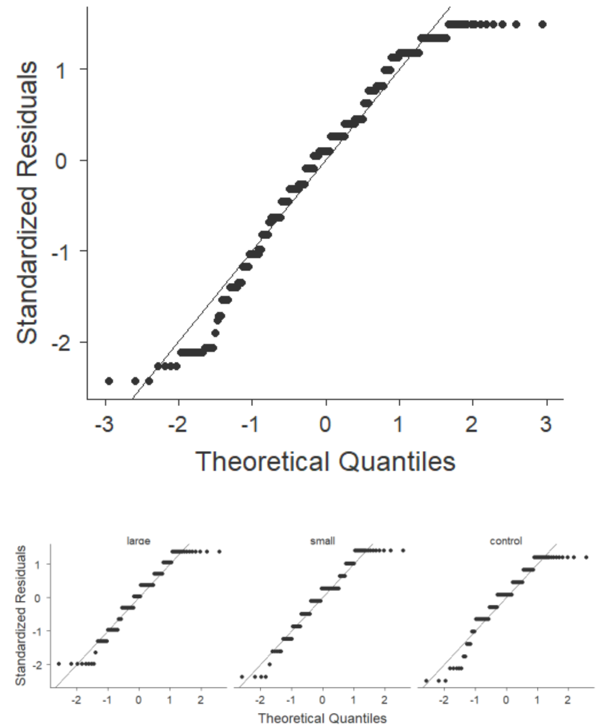
Experiment 4 (Same Condition)

Figure C2. Q-Q Plots of the four samples overall and in large, small, and control conditions (car scenario).

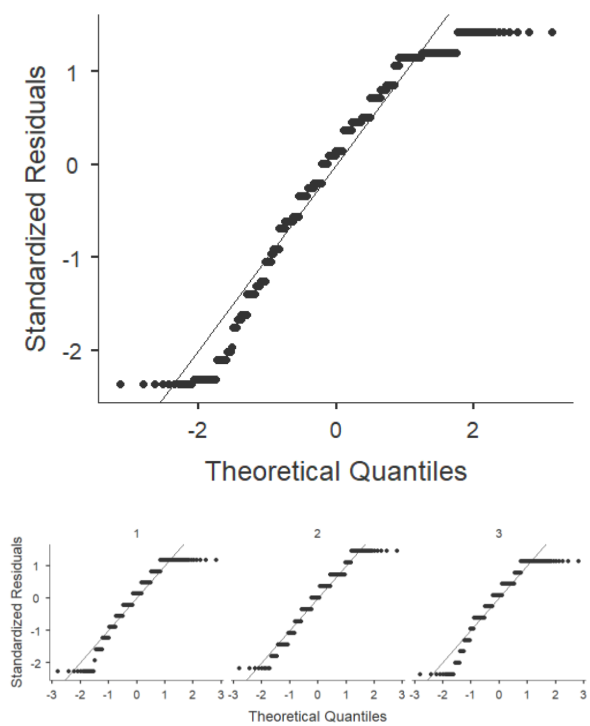
Frequent Flyer Scenario



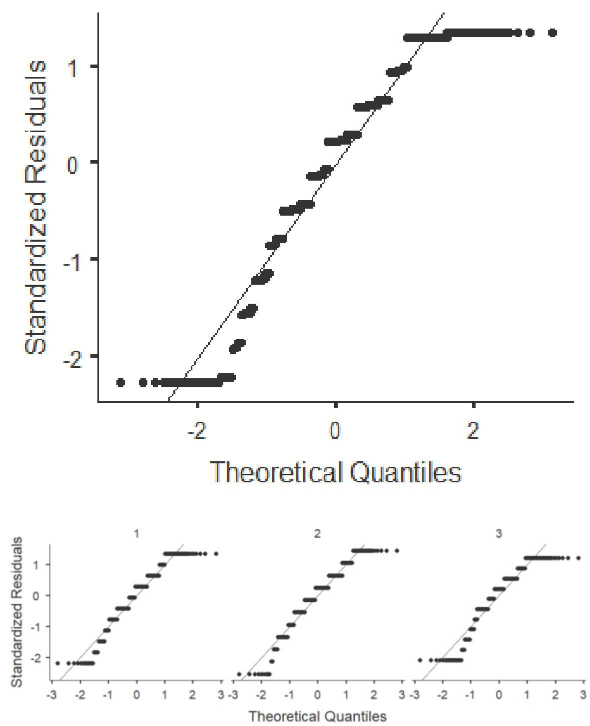
Experiment 1



Experiment 2



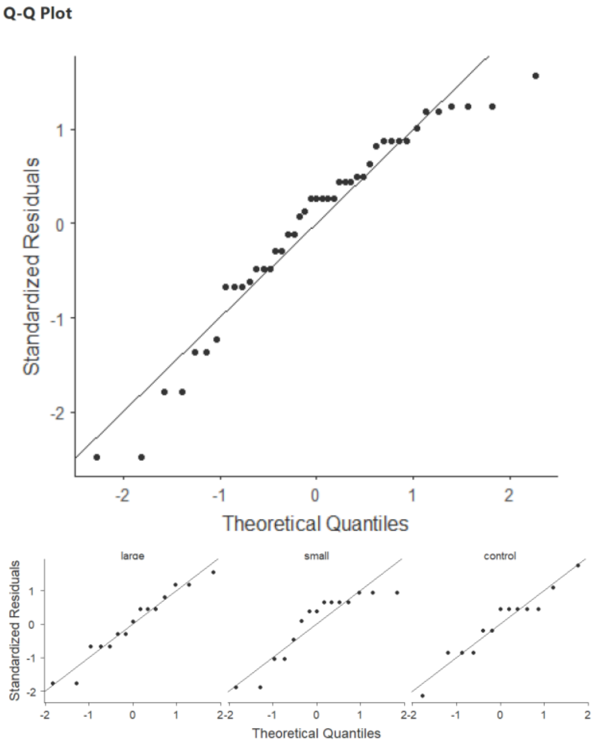
Experiment 3 (Different Conditions)



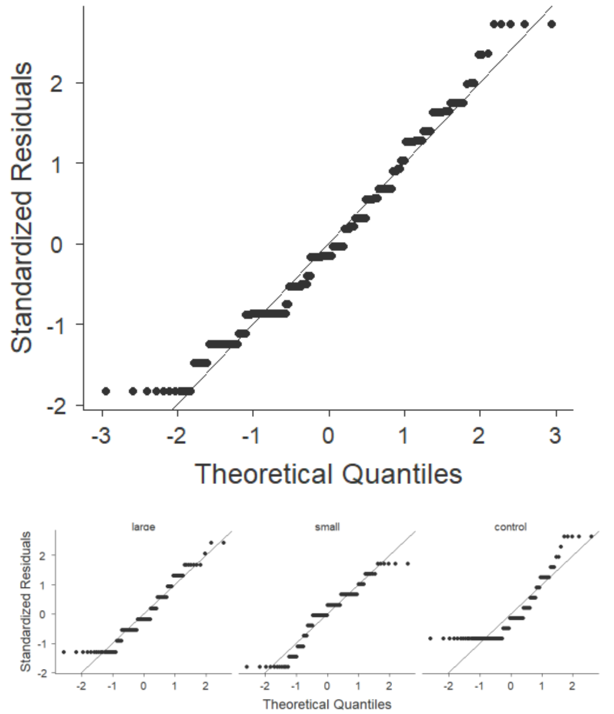
Experiment 4 (Same Condition)

Figure C3. Q-Q Plots of the four samples overall and in large, small, and control conditions (frequent flyer scenario).

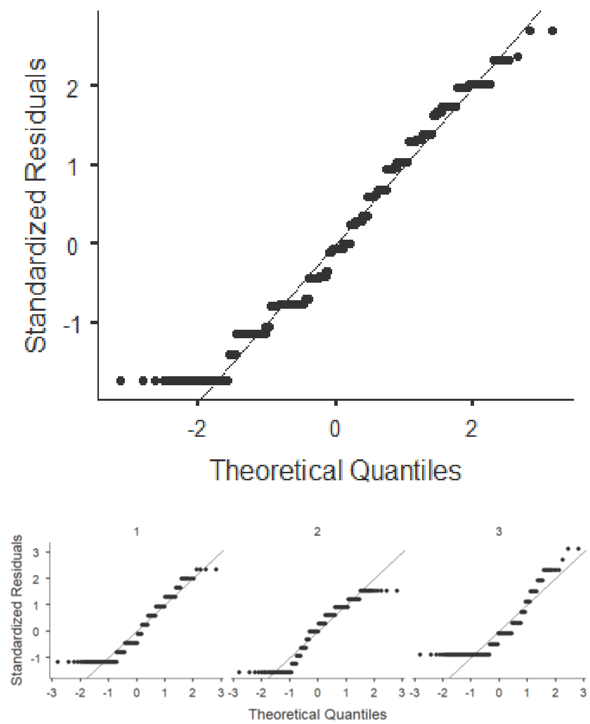
Fitness Center Scenario



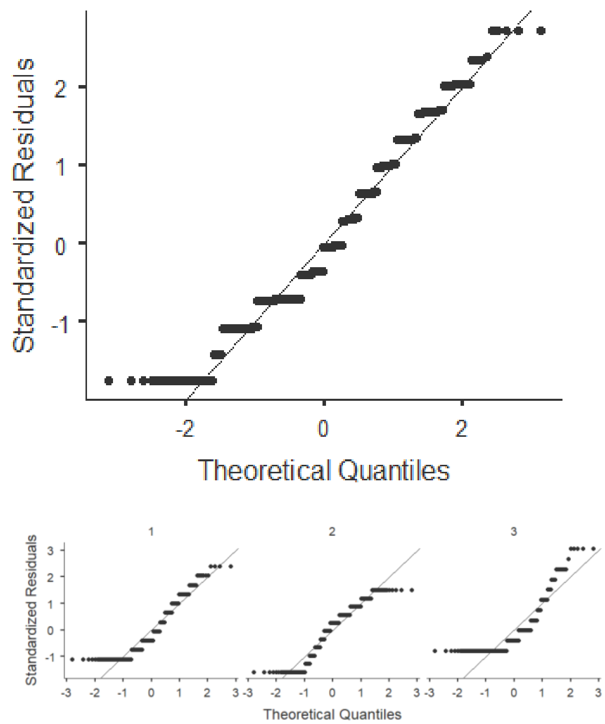
Experiment 1



Experiment 2



Experiment 3 (Different Conditions)



Experiment 4 (Same Condition)

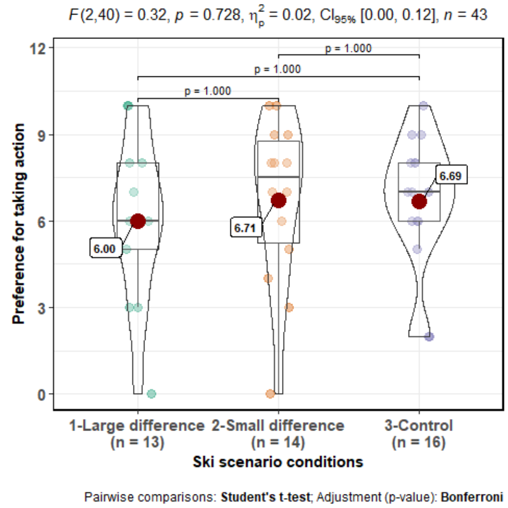
Figure C4. Q-Q Plots of the four samples overall and in large, small, and control conditions (fitness center scenario).

Violin plots

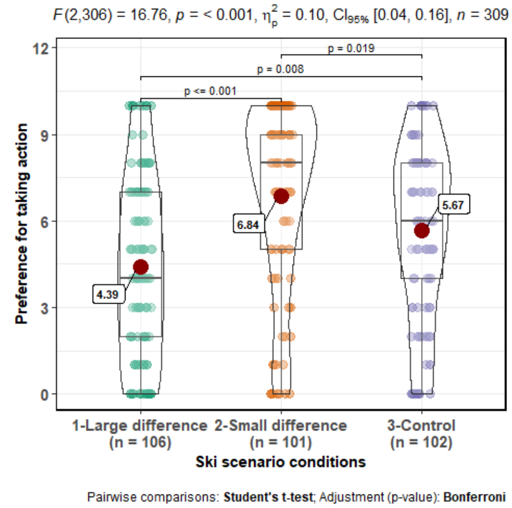
We generated violin plots using the ggbetweenstats commands in R package ggstatsplot (Patil I, 2018). In a violin plot, the red point indicates the mean, the line inside the box indicates the median, the upper and lower boundaries of the box indicate interquartiles, and the upper and lower boundaries of the kernel indicate the 95% confidence interval. Overlapping data points are horizontally jittered to set them apart from each other, thus the width of the kernel represents the frequency of observations (density) at the respective level.

The plots also display p-values using Student's t-test for pairwise comparisons involving groups with equivalent variance, and the Games-Howell test for pairwise comparison when the assumption of equivalence of variance was not met. We also reported partial eta-squared as the effect size measure in order to be consistent with our previous analyses.

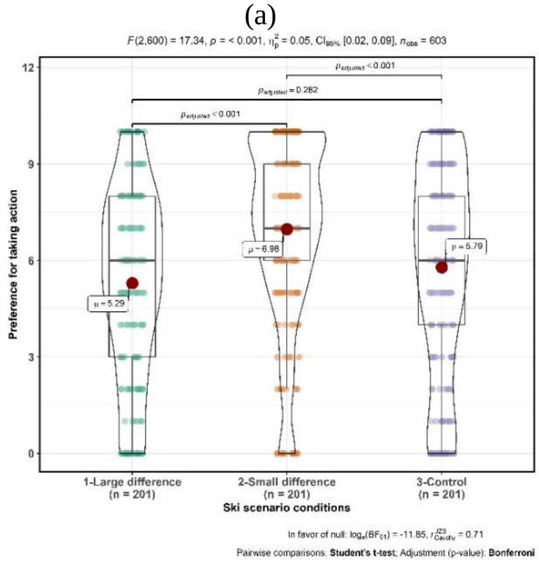
As shown in Figures D1 to D4, the distributions of participants' responses on each condition were highly similar across Experiment 2 and Experiment 3. The distribution of responses in Experiment 1 was slightly similar to the distributions of the other samples in some conditions in some scenarios, but not others. These discrepancies could be due to differences in sample characteristics such as age and cultural background, or to the fact that Experiment 1 has a small sample size. The statistical analysis results shown in the violin plots were all based on Fischer's one-way ANOVA.



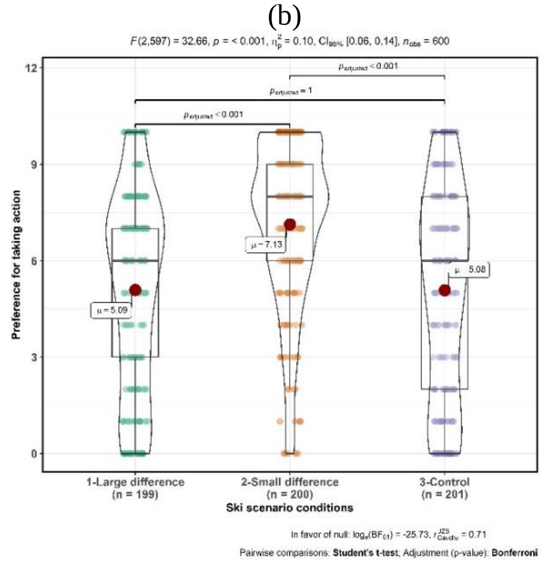
(a)



(b)

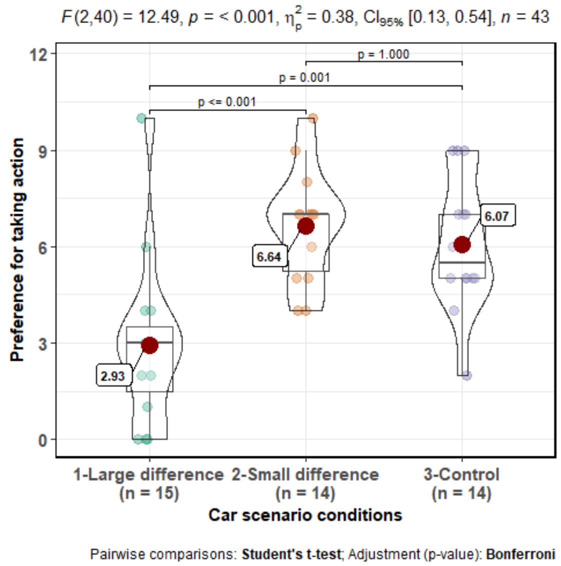


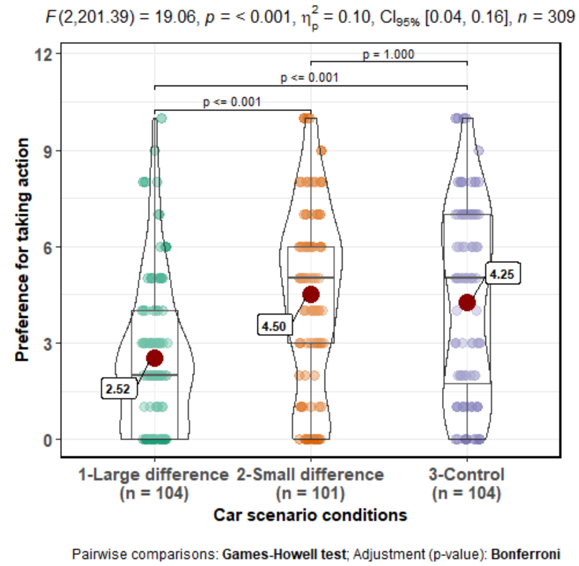
(c)



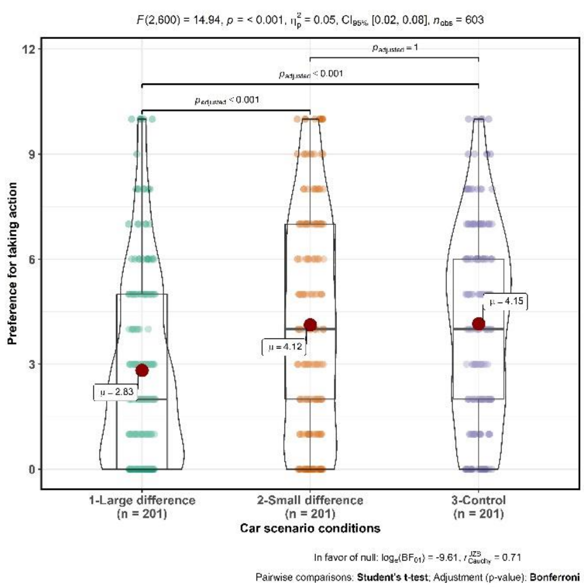
(d)

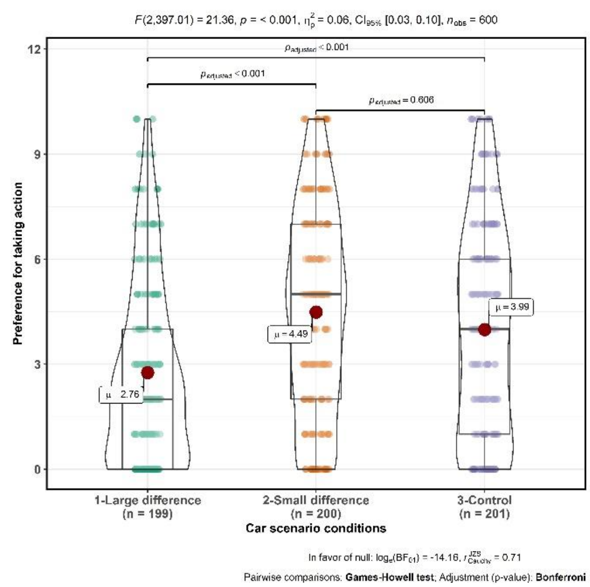
Figure D1. **Ski scenario**: Violin plots of preference for taking action in Experiment 1 (a), Experiment 2 (b), Experiment 3 Different Conditions Design (c), and Experiment 3 Same Condition Design (d).





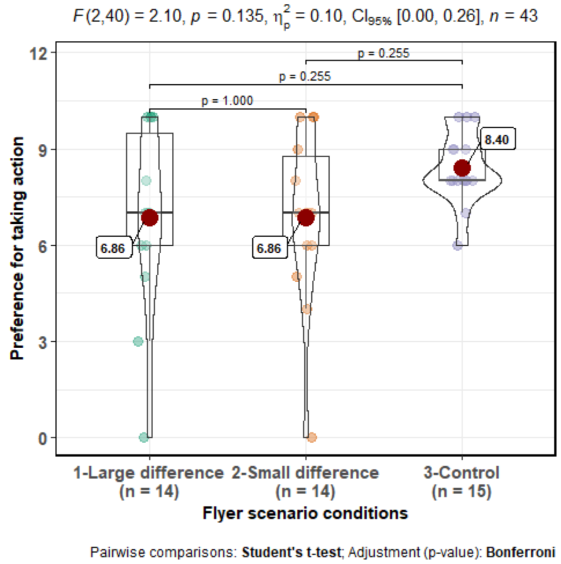
(a) (b)



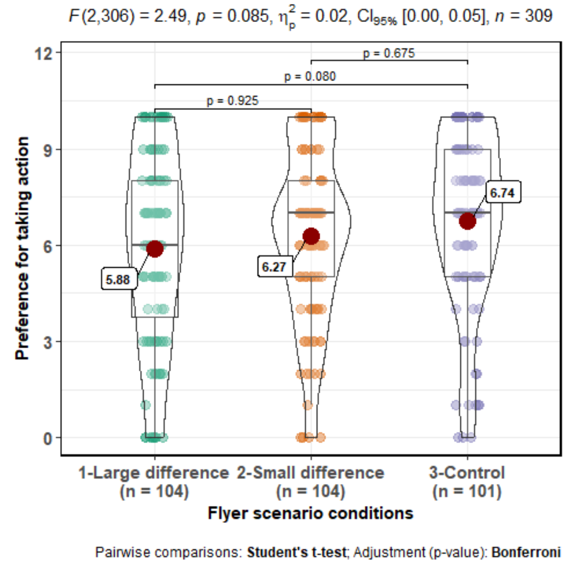


(c) (d)

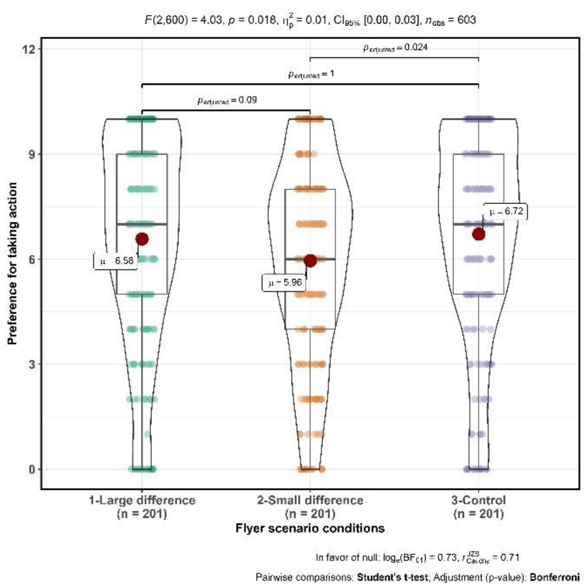
Figure D2. **Car scenario**: Violin plots of preference for taking action in Experiment 1 (a), Experiment 2 (b), Experiment 3 Different Conditions Design (c), and Experiment 3 Same Condition Design (d).



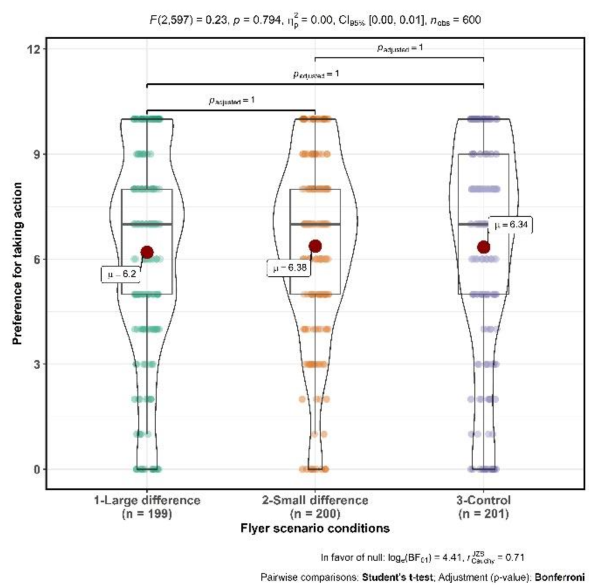
(a)



(b)

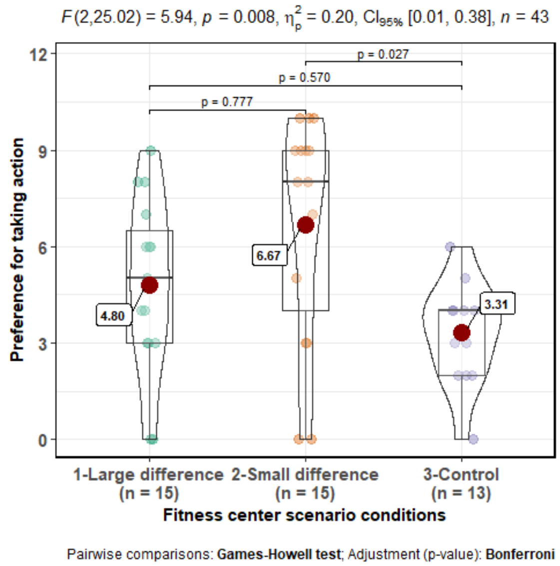


(c)

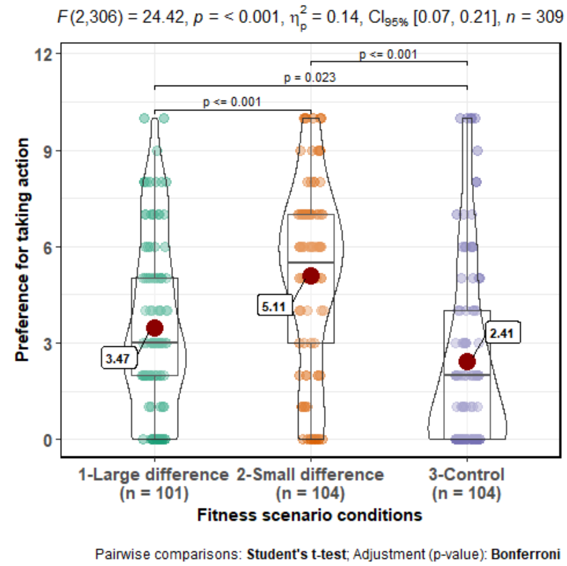


(d)

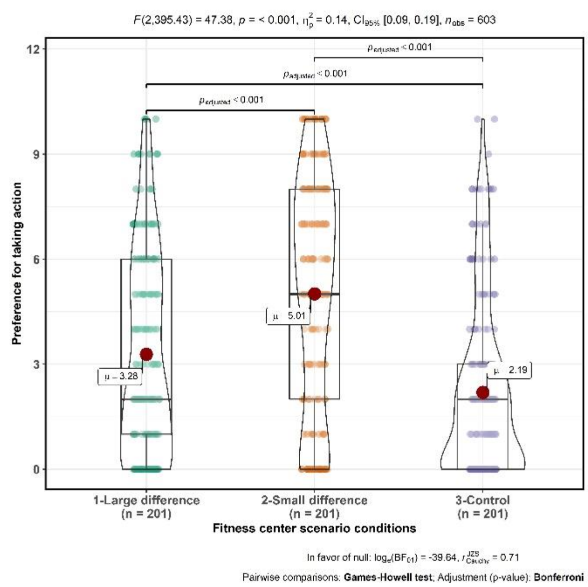
Figure D3. **Frequent flyer scenario**: Violin plots of preference for taking action in Experiment 1 (a), Experiment 2 (b), Experiment 3 Different Conditions Design (c), and Experiment 3 Same Condition Design (d).



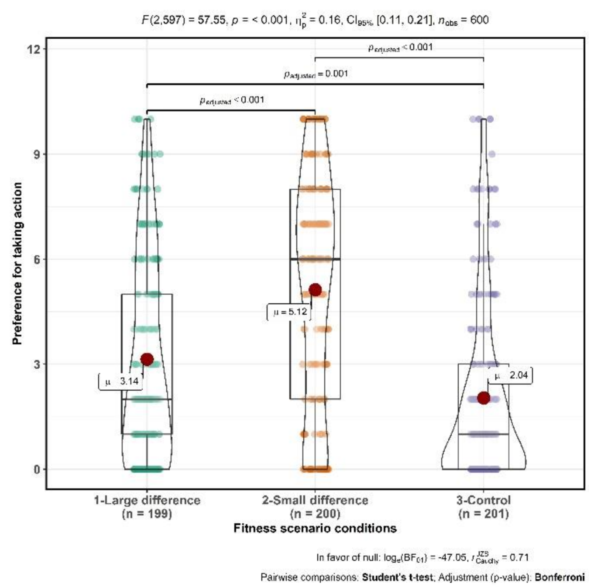
(a)



(b)



(c)



(d)

Figure D4. **Fitness center scenario**: Violin plots of preference for taking action in Experiment 1 (a), Experiment 2 (b), Experiment 3 Different Conditions Design (c), and Experiment 3 Same Condition Design (d).

Results for Welch's ANOVA with Partial Omega-Squared Correction

While Fisher's ANOVA and Student's t-test are conventional options for analysis of variance, these statistical tests can be too conservative (i.e., reduce Type I error) or too sensitive (i.e., increase Type I error) when the assumption of homogeneity of variance is not met (Grissom, 2000; Harwell, Rubinstein, Hayes, & Olds, 1992). Welch's ANOVA and statistical tests that are based on Welch's t-test, such as Games-Howell test are recommended under this type of situations (Delacre, Leys, Mora, & Lakens, 2019; Delacre, Lakens, & Leys, 2017; Wilcox, 1996, p. 135; Keselman, Cribbie, & Holland, 1999; Tomarken & Serlin, 1986). We therefore also present violin plots with results of Welch's ANOVA, Games-Howell test for pairwise comparisons, and omega-squared as the effect size measure in supplementary materials.

<u>Ski Scenario</u>

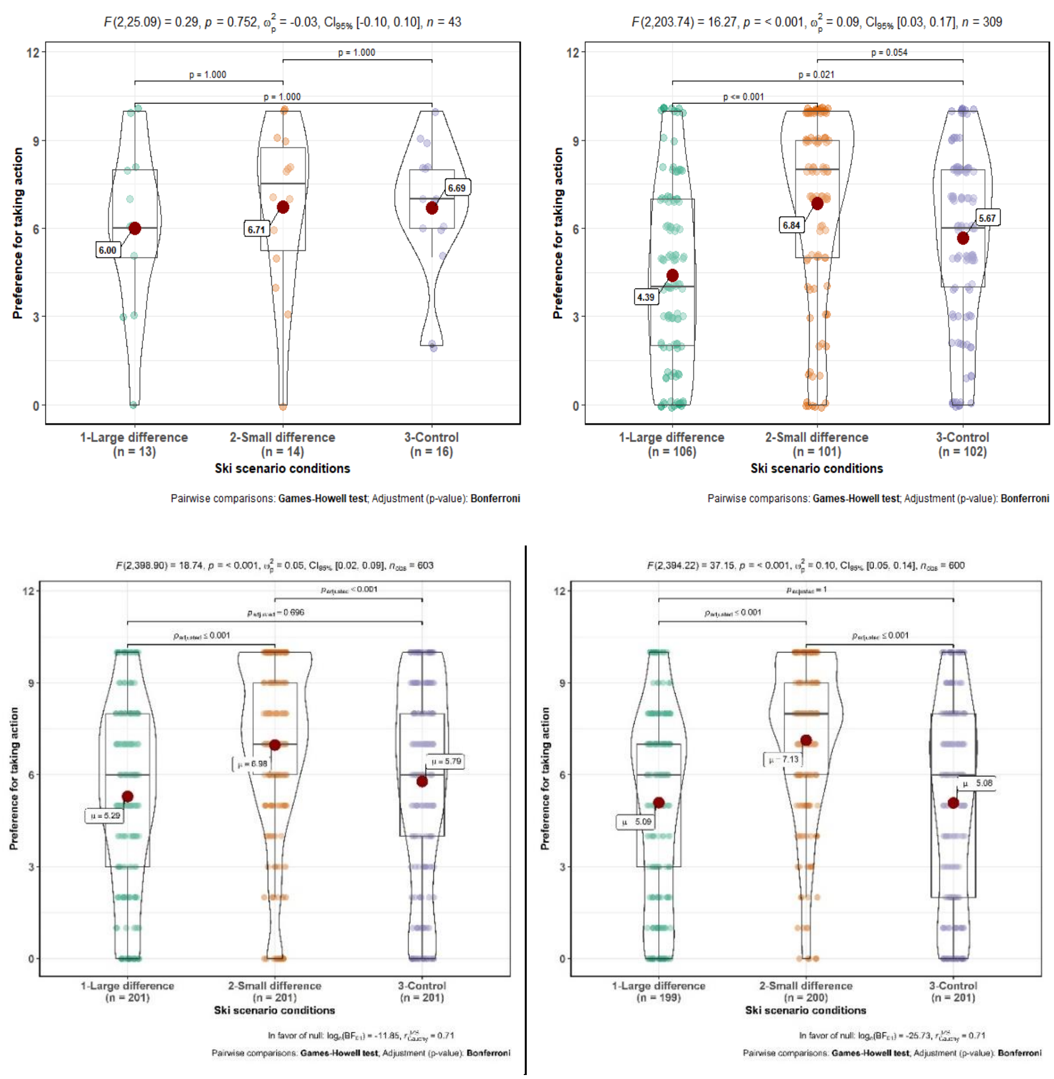


Figure E1. Violin plots of preference for taking action in the ski scenario with Welch's ANOVA, Games-Howell test, and partial omega-squared correction. These plots illustrate the data of the three conditions (large-difference, small-difference, and control) in Experiment 1 (top left panel), Experiment 2 (top right panel), Experiment 3 Different Conditions (bottom left panel), and Experiment 3 Same Condition (bottom right panel).

Car Scenario

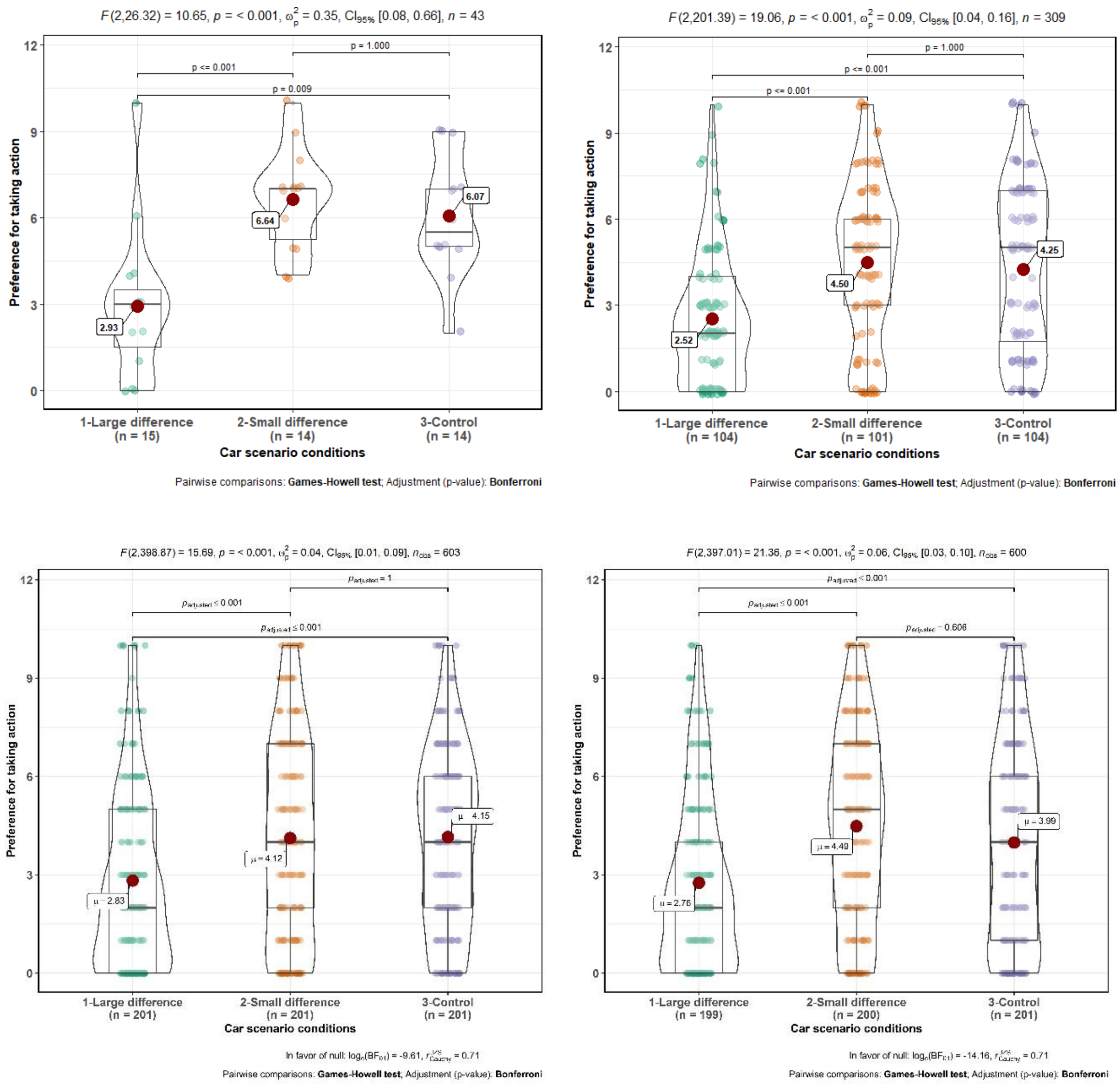


Figure E2. Violin plots of preference for taking action in the car scenario with Welch's ANOVA, Games-Howell test, and partial omega-squared correction. These plots illustrate the data of the three conditions (large-difference, small-difference, and control) in Experiment 1 (top left panel), Experiment 2 (top right panel), Experiment 3 Different Conditions (bottom left panel), and Experiment 3 Same Condition (bottom right panel).

Frequent Flyer Scenario

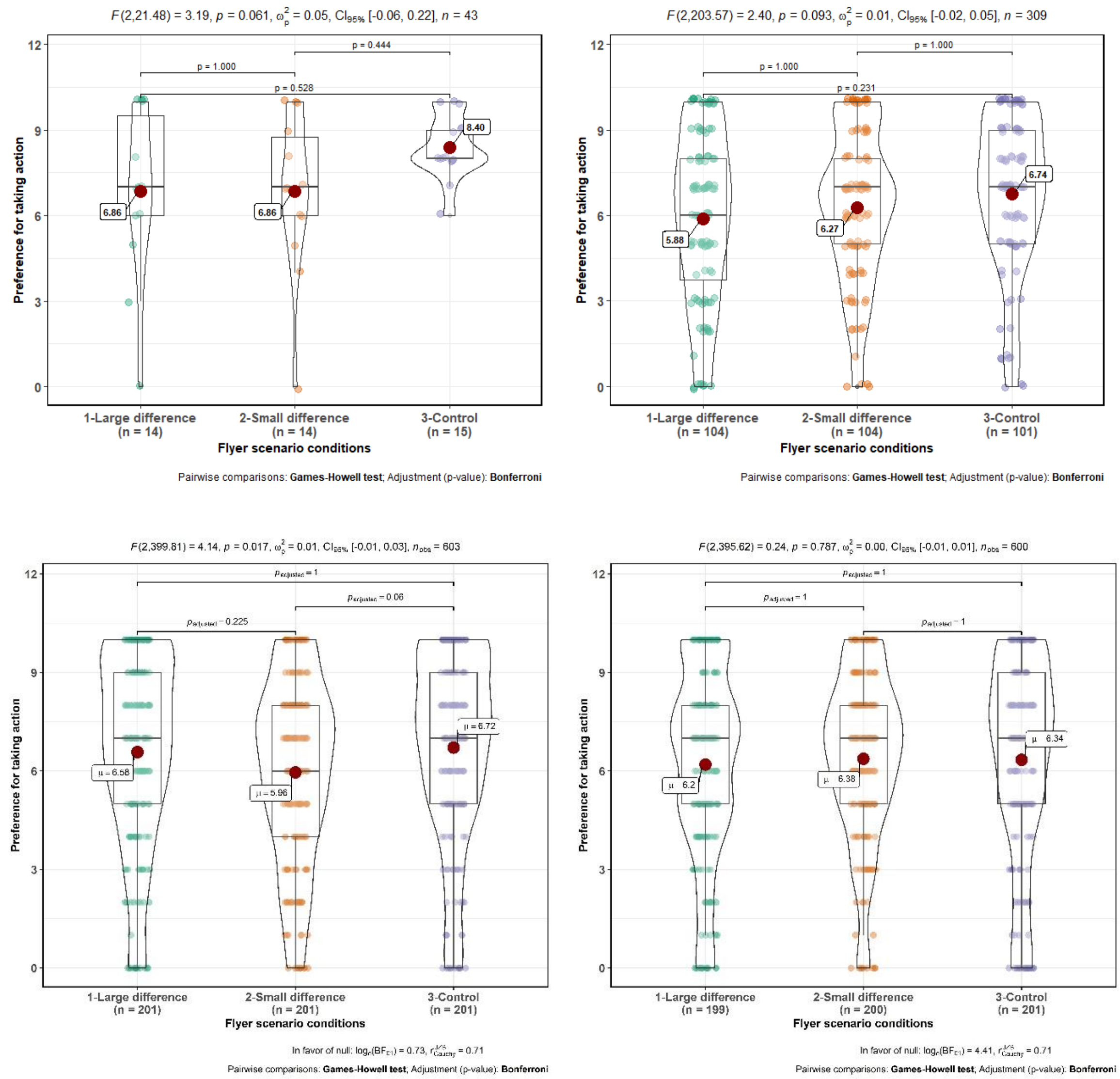


Figure E3. Violin plots of preference for taking action in the frequent flyer scenario with Welch's ANOVA, Games-Howell test, and partial omega-squared correction. These plots illustrate the data of the three conditions (large-difference, small-difference, and control) in Experiment 1 (top left panel), Experiment 2 (top right panel), Experiment 3 Different Conditions (bottom left panel), and Experiment 3 Same Condition (bottom right panel).

Fitness Center Scenario

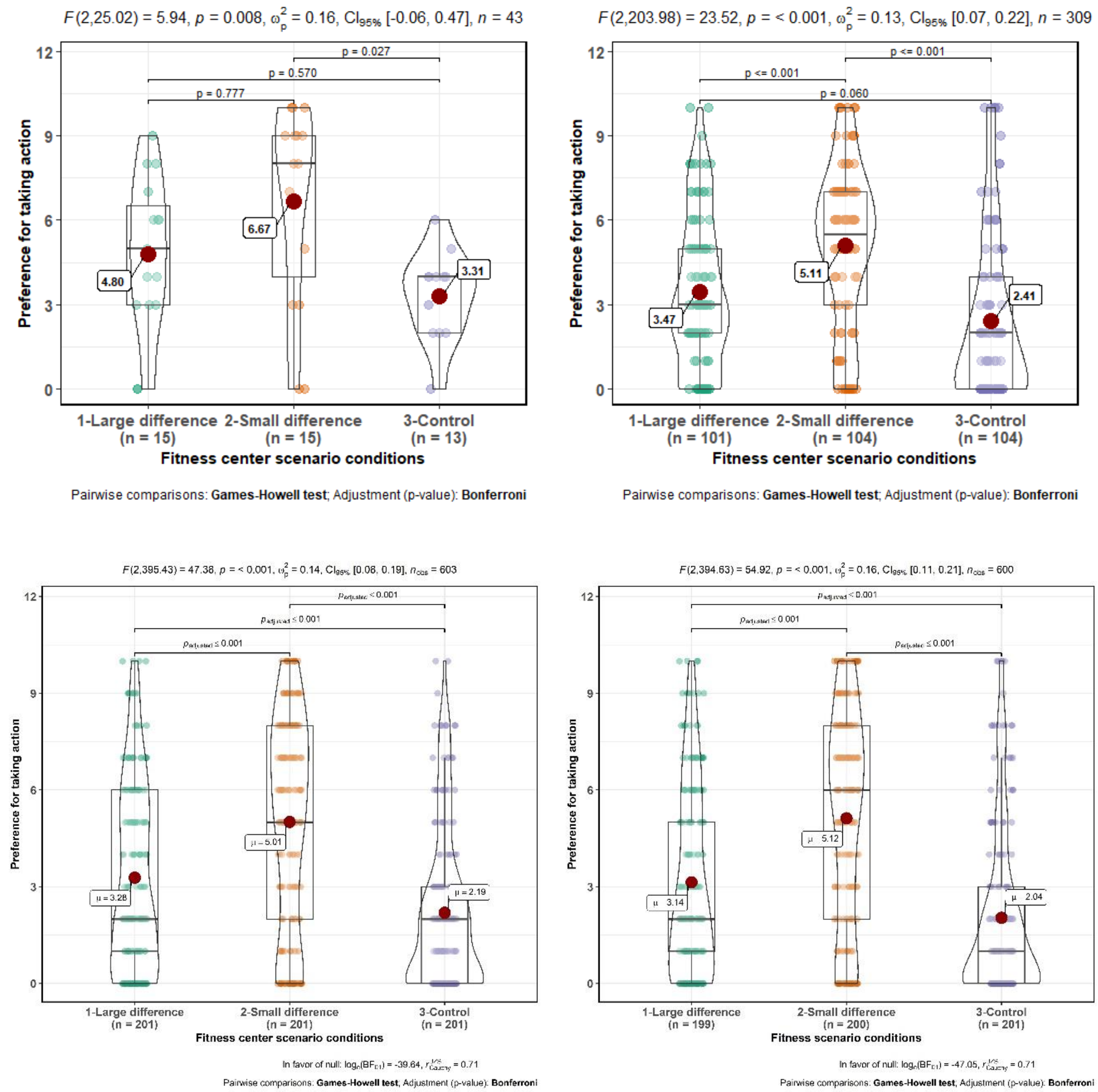


Figure E4. Violin plots of preference for taking action in the fitness center scenario with Welch's ANOVA, Games-Howell test, and partial omega-squared correction. These plots illustrate the data of the three conditions (large-difference, small-difference, and control) in Experiment 1 (top left panel), Experiment 2 (top right panel), Experiment 3 Different Conditions (bottom left panel), and Experiment 3 Same Condition (bottom right panel).

Minor deviations from pre-registration plans

There are a few points that we had deviated from the pre-registration. First, while the students wrote that they would conduct an ANOVA or a mixed ANOVA for the analysis that combined data across all four scenarios, later on we chose mixed-effects regression, which is a better strategy to analyze the data. Second, regarding the seriousness question, while the students proposed a cut-off point of five on a seven-point Likert scale for exclusion, in practice we used a five-point Likert scale for this question. Therefore, in robustness checks, we adapted the cut-off point (from 5 out of 7 to 4 out of 5) for exclusion based on the seriousness question. Third, students' pre-registration documents for Experiment 2 did not specify any test of statistical assumptions (e.g., equal number of participants in each condition, equivalence of variance, and normality of residuals). We still conducted these tests for Experiment 2 so that the analyses are comparable to those for Experiment 1.

Gender as a covariate (ANCOVA)

While Tykocinski et al. (1995) described their analyses in the main text as ANOVA, they mentioned in the footnote of page 795 that “In all of the six experiments reported in this article, we conducted analyses including sex of participant as a variable.” The main effect of the experimental conditions (i.e., large-difference, small-difference, and control conditions) remained the same, and there were a few differences in pairwise comparisons. In terms of the main effect of gender, Tykocinski et al. (1995) reported that gender only had a significant main effect on the dependent variable (likelihood) in Scenario 1, but not in Scenarios 2, 3, and 4; in our studies, gender only had a marginally significant main effect in Experiment 1 Scenario 4 (F (1, 39) = 3.16, p = .08) and Experiment 2 Scenario 2 (F (1, 305) = 6.60, p = .01), but not in other scenarios.

We note a deviation from pre-registration for the exploratory from ANOVA to ANCOVA which is the appropriate approach to handle situations where gender is a covariate.

The main effect of the experimental conditions (i.e., large-difference, small-difference, and control conditions) remained the same in all four samples. Pairwise comparisons also showed largely consistent results. For pairwise comparisons, we reported both the p values based on both the more lenient LSD method and Tukey’s test. Interested readers can obtain the p values for the Bonferroni method by dividing the p value reported using the LSD method by three.

In terms of the main effect of gender, Tykocinski et al. (1995) reported that gender only had a significant main effect on the dependent variable (likelihood) in Scenario 1, but not in Scenarios 2, 3, and 4. In our studies, gender only had a marginally significant main effect in Experiment 1 Scenario 4 (F (1, 39) = 3.16, p = .08), Experiment 2 Scenario 2 (F (1, 305) = 6.60, p = .01), Experiment 3 (Different Conditions Design) Scenario 2 (F (1, 599) = 4.55, p = .03), Scenario 4 (F (1, 599) = 11.02, p < .001). It was not a significant predictor of likelihood of acting on the subsequent opportunity in any other scenarios or analyses.

Table G1

One-way ANCOVA on the Mean Likelihood of Acting on the Subsequent Opportunity (Experiment 1, n = 43)

Scenario	Conditions			*df*	*F*	*p*	*f*	η^2
	Large-difference	Small-difference	Control					
Ski	6.11	6.79	6.53	2, 39	.23	.79	0.11	0.01
Car	2.94^a	6.65^b	6.06^b	2, 39	12.11***	< .001	0.79	0.38
Frequent flyer	6.87	6.69	8.54	2, 39	2.54	.09	0.11	0.01
Fitness center	4.52^a	6.95^b	3.30^a	2, 39	6.37**	.004	0.55	0.23

Note. Because we controlled for gender as a covariate, estimated marginal means rather than straight means in descriptive analyses were reported. Higher numbers indicate higher likelihoods of taking action on a subsequent opportunity on an 11-point scale. * p < .05. ** p < .01, *** p < .001.

Table G2

Pairwise Comparisons of Mean Likelihood of Acting on the Subsequent Opportunity following One-Way ANCOVA (Experiment 1, n = 43)

Comparison	*t*	Cohen's *d*	95% CI	*p*	p_{Tukey}
Ski					
large vs. small	-.68	-0.26	[-1.02, 0.50]	.50	.78
large vs. control	-.43	-0.16	[-0.89, 0.57]	.67	.90
small vs. control	.26	0.10	[-0.62, 0.81]	.79	.96
Car					
large vs. small	-4.56***	-1.69	[-2.54, -0.83]	< .001	< .001
large vs. control	-3.83**	-1.42	[-2.23, -0.59]	< .001	.001
small vs. control	.70	0.27	[-0.48, 1.01]	.49	.76
Frequent flyer					
large vs. small	.20	0.07	[-0.67, 0.82]	.84	.98
large vs. control	-1.89	-0.70	[-1.45, 0.05]	.07	.15
small vs. control	-2.00	-0.74	[-1.49, 0.02]	.05	.13
Fitness center					
large vs. small	-2.33*	-0.85	[-1.59, -0.09]	.03	.06
large vs. control	1.18	0.45	[-0.31, 1.2]	.25	.47
small vs. control	3.51**	1.33	[0.49, 2.14]	.00	.00

Note. Large = large-difference condition; Small = small-difference condition; Control = control condition. The marking of the asterisks are based on unadjusted *p* value calculated by Fisher's LSD method. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table G3

One-way ANCOVA on the Mean Likelihood of Acting on the Subsequent Opportunity (Experiment 2, n = 309)

Scenario	Conditions			*df*	*F*	*p*	*f*	η^2
	Large-difference	Small-difference	Control					
Ski	4.39[a]	6.84[b]	5.67[c]	2, 305	16.71***	< .001	0.33	0.10
Car	2.49[a]	4.49[b]	4.29[b]	2, 305	17.68***	< .001	0.34	0.10
Frequent flyer	5.89	6.26	6.73	2, 305	2.34	.10	0.12	0.02
Fitness center	3.47[a]	5.12[b]	2.40[c]	2, 305	24.76***	< .001	0.40	0.14

Note. Because we controlled for gender as a covariate, estimated marginal means rather than straight means in descriptive analyses were reported. Higher numbers indicate higher likelihoods of taking action on a subsequent opportunity on an 11-point scale. * $p < .05$. ** $p < .01$, *** $p < .001$.

Table G4

Pairwise Comparisons of Mean Likelihood of Acting on the Subsequent Opportunity following One-Way ANCOVA (Experiment 2, n = 309)

Comparison	*t*	Cohen's *d*	95% CI	*p*	p_{Tukey}
Ski					
large vs. small	-5.78***	-0.80	[-1.09, -0.52]	< .001	< .001
large vs. control	-3.01**	-0.42	[-0.69, -0.14]	.00	.01
small vs. control	2.73*	0.38	[0.11, 0.66]	.01	.02
Car					
large vs. small	-5.39***	-0.75	[-1.03, -0.47]	< .001	< .001
large vs. control	-4.87***	-0.67	[-0.95, -0.39]	< .001	< .001
small vs. control	0.54	0.08	[-0.2, 0.35]	.59	.85
Frequent flyer					
large vs. small	-0.97	-0.13	[-0.41, 0.14]	.34	.60
large vs. control	-2.16	-0.30	[-0.58, -0.03]	.03	.08
small vs. control	-1.20	-0.17	[-0.44, 0.11]	.23	.45
Fitness center					
large vs. small	-4.21***	-0.59	[-0.87, -0.31]	< .001	< .001
large vs. control	2.73*	0.38	[0.1, 0.66]	.01	.02
small vs. control	6.98***	0.97	[0.68, 1.25]	< .001	< .001

Note. Large = large-difference condition; Small = small-difference condition; Control = control condition. The marking of the asterisks are based on unadjusted *p* value calculated by Fisher's LSD method. Because the main effect in the frequent flyer condition was not significant, we do not mark the pairwise comparisons here. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table G5

One-way ANCOVA on the Mean Likelihood of Acting on the Subsequent Opportunity (Experiment 3 Different Conditions Design, n = 603)

Scenario	Conditions							
	Large-difference	Small-difference	Control	*df*	*F*	*p*	*f*	η^2
Ski	5.29[a]	6.98[b]	5.79[a]	2, 599	17.41***	< .001	0.23	0.05
Car	2.84[a]	4.13[b]	4.14[b]	2, 599	14.53***	< .001	0.23	0.05
Frequent flyer	6.57[a,b]	5.97[a]	6.72[b]	2, 599	3.89	.02	0.11	0.01
Fitness center	3.31[a]	4.99[b]	2.19[c]	2, 599	48.92***	< .001	0.40	0.14

Note. Because we controlled for gender as a covariate, estimated marginal means rather than straight means in descriptive analyses were reported. Higher numbers indicate higher likelihoods of taking action on a subsequent opportunity on an 11-point scale. * $p < .05$. ** $p < .01$, *** $p < .001$.

Table G6

Pairwise Comparisons of Mean Likelihood of Acting on the Subsequent Opportunity following One-Way ANCOVA (Experiment 3 Different Conditions Design, n = 603)

Comparison	*t*	Cohen's *d*	95% CI	*p*	p_{Tukey}
Ski					
large vs. small	-5.74***	-0.57	[-0.77, -0.37]	< .001	< .001
large vs. control	-1.68	-0.17	[-0.36, 0.03]	.09	.21
small vs. control	4.06***	0.40	[0.21, 0.60]	< .001	< .001
Car					
large vs. small	-4.65***	-0.46	[-0.66, -0.27]	< .001	< .001
large vs. control	-4.69***	-0.47	[-0.67, -.27]	< .001	< .001
small vs. control	-0.05	-0.005	[-0.20, 0.19]	.96	1.00
Frequent flyer					
large vs. small	2.12	0.21	[0.02, 0.41]	.04	.09
large vs. control	-0.52	-0.05	[-0.24, 0.14]	.61	.86
small vs. control	-2.63*	-0.26	[-0.46, -0.07]	.01	.02
Fitness center					
large vs. small	-5.91***	-0.59	[-0.79, -0.39]	< .001	< .001
large vs. control	3.91***	0.39	[0.19, 0.59]	< .001	< .001
small vs. control	9.82***	0.98	[0.77, 1.19]	< .001	< .001

Note. Large = large-difference condition; Small = small-difference condition; Control = control condition. The marking of the asterisks are based on unadjusted *p* value calculated by Fisher's LSD method. Because the main effect in the frequent flyer condition was not significant, we do not mark the pairwise comparisons here. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table G7

Results of Mixed Factorial ANCOVA with Gender as a Covariate for Experiment 3 (Same Condition Design) (N = 600)

	Sum of Squares	*df*	*F*	*p*	f_G	η^2_G	f_p	η^2_p
(Within-subject)								
Scenario	245.61	3	13.28	< .001	.11	.01	.14	.02
Scenario * Condition	726.34	6	19.64	< .001	.19	.04	.25	.06
Scenario * Gender	29.94	3	1.62	.18	.04	.002	.05	.001
$\text{Residual}_{\text{(Within)}}$	11021.38	1788						
(Between-subject)								
Condition	1121.36	2	38.92	< .001	.24	.05	.37	.12
Gender	0.05	1	0.00	0.96	.00	0.00	.00	0.00
$\text{Residual}_{\text{(Between)}}$	8586.74	596						

Table G8

Pairwise Comparisons of Mixed Factorial ANCOVA with Gender as a Covariate for Experiment 3 (Same Condition Design) (N = 600)

	Mean Difference	*t*	P_{Tukey}	Cohen's *d*	95% CI for Cohen's *d*
Across scenarios					
large vs. small	-1.48	-7.79	< .001	-0.78	[-0.98, -0.58]
large vs. control	-0.07	-0.35	0.93	-0.04	[-0.23, 0.16]
small vs. control	1.42	7.47	< .001	0.75	[0.54, 0.95]
Ski					
large vs. small	-2.03	-7.07	< .001	-0.71	[-0.91, -0.51]
large vs. control	0.01	0.05	1.00	0.01	[-0.19, 0.20]
small vs. control	2.05	7.15	< .001	0.71	[0.51, 0.92]
Car					
large vs. small	-1.73	-6.00	< .001	-0.60	[-0.80, -0.40]
large vs. control	-1.22	-4.25	.001	-0.43	[-0.62, -0.23]
small vs. control	0.50	1.75	.84	0.17	[-0.02, 0.37]
Frequent flyer					
large vs. small	-0.16	-0.57	1.00	-0.06	[-0.25, 0.14]
large vs. control	-0.13	-0.44	1.00	-0.04	[-0.24, 0.15]
small vs. control	0.04	0.13	1.00	0.01	[-0.18, 0.21]
Fitness center					
large vs. small	-2.01	-6.99	< .001	-0.70	[-0.90, -0.50]
large vs. control	1.07	3.71	0.01	0.37	[0.17, 0.57]
small vs. control	3.08	10.73	< .001	1.07	[0.86, 1.28]

Note. The *p* values of pairwise comparisons were adjusted using the Tukey's honestly significant difference method.

Experiment 2 results with exclusions

In Tables H1 and H2, we present the results analyzed based on the sample of participants who met the exclusion criteria based on English proficiency (i.e., scoring no less than 5 out of 7 on the question: How would you generally rate your understanding of the English used in this study)?

Table H1

One-way ANOVA on the Mean Likelihood of Acting on the Subsequent Opportunity (After Exclusion, Experiment 2, n = 305)

	Conditions							
Scenario	Large-difference	Small-difference	Control	*df*	*F*	*p*	*f*	η^2
Ski	4.37[a]	6.97[b]	5.67[c]	2, 302	18.9***	<.001	0.35	0.11
Car	2.52[a]	4.54[b]	4.19[b]	2, 302	19.1***	<.001	0.33	0.10
Frequent flyer	5.88	6.23	6.8	2, 302	2.86	.059	0.14	0.02
Fitness center	3.46[a]	5.11[b]	2.41[c]	2, 302	24.2***	<.001	0.40	0.14

Note. Higher numbers indicate higher likelihoods of taking action on a subsequent opportunity on an 11-point scale. * $p < .05$. ** $p < .01$, *** $p < .001$.

Table H2

Pairwise Comparisons of Mean Likelihood of Acting on the Subsequent Opportunity (After Exclusion, Experiment 2, n = 305)

Comparison	*t*	Cohen's *d*	95% CI	*p*
Ski				
large vs. small	-6.14***	-0.86	[-1.15, -0.57]	<.001
large vs. control	-3.09**	-0.43	[-0.71, -0.15]	.01
small vs. control	3.06**	0.43	[0.15, 0.71]	.01
Car				
large vs. small	-5.39***	-0.75	[-1.04, -0.47]	<.001
large vs. control	-4.46***	-0.62	[-0.90, -0.34]	<.001
small vs. control	0.93	0.13	[-0.15, 0.41]	.66
Frequent flyer				
large vs. small	-0.91	-0.13	[-0.40, 0.15]	.64
large vs. control	-2.38	-0.33	[-0.61, -0.06]	.05
small vs. control	-1.47	-0.21	[-0.48, 0.07]	.31
Fitness center				
large vs. small	-4.20***	-0.59	[-0.87, -0.31]	< .001
large vs. control	2.67*	0.38	[0.10, 0.66]	.02
small vs. control	6.90***	0.96	[0.67, 1.25]	< .001

Note. Large = large-difference condition; Small = small-difference condition; Control = control condition. We used Tukey's test for Scenarios 1, 3, and 4, and Games-Howell's test for Scenario 2. * $p < .05$. ** $p < .01$. *** $p < .001$.

In Tables H3 and H4, we present the results analyzed based on the sample of participants who met the exclusion criteria based on both English proficiency (i.e., scoring no less than 5 out of 7 on the question: How would you generally rate your understanding of the English used in this study) and seriousness (i.e., scoring no less than 4 out of 5 on the question: "How serious are you when filling the survey?").

Table H3

One-way ANOVA on the Mean Likelihood of Acting on the Subsequent Opportunity (After Exclusion, Experiment 2, n = 299)

	Conditions							
Scenario	Large-difference	Small-difference	Control	*df*	*F*	*p*	*f*	η^2
Ski	4.42[a]	6.97[b]	5.68[c]	2, 296	17.4***	<.001	0.34	0.11
Car	2.49[a]	4.54[b]	4.12[b]	2, 296	16.5***	<.001	0.33	0.10
Frequent flyer	5.91[a]	6.28[a, b]	6.90[b]	2, 296	3.25*	.04	0.15	0.02
Fitness center	3.47[a]	5.08[b]	2.34[c]	2, 296	23.9***	<.001	0.40	0.14

Note. Higher numbers indicate higher likelihoods of taking action on a subsequent opportunity on an 11-point scale. * $p < .05$. ** $p < .01$, *** $p < .001$.

Table H4

Pairwise Comparisons of Mean Likelihood of Acting on the Subsequent Opportunity (After Exclusion, Experiment 2, n = 299)

Comparison	*t*	Cohen's *d*	95% CI	*p*
Ski				
large vs. small	-5.90***	-0.84	[-1.13, -0.55]	<.001
large vs. control	-2.96**	-0.42	[-0.69, -0.14]	.01
small vs. control	2.97**	0.42	[0.14, 0.71]	.01
Car				
large vs. small	-5.43***	-0.76	[-1.05, -0.48]	<.001
large vs. control	-4.29***	-0.61	[-0.89, -0.32]	<.001
small vs. control	1.09	0.16	[-0.12, 0.43]	.56
Frequent flyer				
large vs. small	-0.96	-0.13	[-0.41, 0.14]	.61
large vs. control	-2.53*	-0.36	[-0.64, -0.08]	.03
small vs. control	-1.57	-0.22	[-0.50, 0.06]	.26
Fitness center				
large vs. small	-4.04***	-0.57	[-0.85, -0.29]	<.001
large vs. control	2.83*	0.40	[0.12, 0.68]	.01
small vs. control	6.88***	0.97	[0.68, 1.27]	<.001

Note. Large = large-difference condition; Small = small-difference condition; Control = control condition. We used Tukey's test for Scenarios 1, 3, and 4, and Games-Howell's test for Scenario 2. * $p < .05$. ** $p < .01$. *** $p < .001$.

Experiment 3 results with exclusions

In Tables J1 through J4, we present the results analyzed based on the sample of participants who met the exclusion criteria based on the criteria specified in the pre-registration.

Specific criteria:

- Participants who have seen the study materials (or similar materials) before will be removed from the sample.

General criteria:

- All participants indicating a low proficiency of English (self-report < 5, on a 1-7 scale)
- Participants who report not being serious about filling in the survey (self-report < 4, on a 1-5 scale).
- Participants who completed the survey too quickly (within one minute).

The sample size for Experiment 3 (Different Conditions Design) after exclusion is 595 (199 in large-difference condition, 198 in small-difference condition, and 198 in control condition). The results are largely the same as those reported in the main text, except that the main effect in the frequent flyer condition became significant (F (2.592) = 3.96, p = .02), with pairwise comparison showing that the likelihood to act on subsequent opportunity was higher in small-difference condition than in control condition (t = -2.62, p = .03).

Table J1

One-way ANOVA on the Mean Likelihood of Acting on the Subsequent Opportunity (After Exclusion, Experiment 3 Different Conditions Design, n = 595)

	Conditions							
Scenario	Large-difference	Small-difference	Control	*df*	*F*	*p*	*f*	η^2
Ski	5.30^a	7.04^b	5.76^a	2, 592	18.75***	<.001	0.25	0.06
Car	2.78^a	4.13^b	4.15^b	2, 592	15.85***	<.001	0.23	0.05
Frequent flyer	$6.60^{a,b}$	5.96^a	6.71^b	2, 592	3.96	.02	0.10	0.01
Fitness center	3.25^a	5.04^b	2.20^c	2, 592	48.98***	<.001	0.40	0.14

Note. Higher numbers indicate higher likelihoods of taking action on a subsequent opportunity on an 11-point scale. * p < .05. ** p < .01, *** p < .001.

Table J2

Pairwise Comparisons of Mean Likelihood of Acting on the Subsequent Opportunity (After Exclusion, Experiment 3 Different Conditions Design, n = 595)

Comparison	*t*	Cohen's *d*	95% CI	*p*
Ski				
large vs. small	-6.19***	-0.62	[-0.82, -0.42]	<.001
large vs. control	-1.54	-0.15	[-0.35, 0.04]	.27
small vs. control	4.27***	0.43	[0.23, 0.63]	<.001
Car				
large vs. small	-4.75***	-0.48	[-0.68, -0.28]	<.001
large vs. control	-5.13***	-0.51	[-0.71, -0.31]	<.001
small vs. control	-0.07	-0.01	[-0.20, 0.19]	1.00
Frequent flyer				
large vs. small	2.20	0.22	[0.02, 0.42]	.07
large vs. control	-0.41	-0.04	[-0.24, 0.16]	.91
small vs. control	-2.62*	-0.26	[-0.46, -0.07]	.03
Fitness center				
large vs. small	-5.83***	-0.59	[-0.79, -0.38]	< .001
large vs. control	3.89***	0.39	[0.19, 0.59]	< .001
small vs. control	9.72***	0.98	[0.77, 1.18]	< .001

Note. Large = large-difference condition; Small = small-difference condition; Control = control condition. We used Tukey's test for Scenario 3, and Games-Howell's test for the other three scenarios. * $p < .05$. ** $p < .01$. *** $p < .001$.

The sample size for Experiment 3 (Same Condition Design) after exclusion is 594 (198 in large-difference condition, 195 in small-difference condition, and 201 in control condition). The results are largely the same as those reported in the main text.

Table J3

Results of Mixed Factorial ANOVA (After Exclusion, Experiment 3 Same Condition Design, n = 594)

	Sum of Squares	*df*	*F*	*p*	f_G	f_P	η^2_G	η^2_P
(Within-subject)								
Scenario	3723.31	3	200.79	< .001	.44	.58	.16	.25
Scenario * Condition	708.41	6	19.10	< .001	.20	.25	.04	.06
$\text{Residual}_{(\text{Within})}$	10959.14	1773						
(Between-subject)								
Condition	1158.73	2	40.69	< .001	.25	.37	.06	.12
$\text{Residual}_{(\text{Between})}$	8414.14	591						

Table J4

Pairwise Comparisons following Mixed Factorial ANOVA (After Exclusion, Experiment 3 Same Condition Design, n = 594)

	Mean Difference	*t*	P_{Tukey}	Cohen's *d*	95% CI for Cohen's *d*
Across scenarios					
large vs. small	-1.52	-8.00	< .001	-0.81	[-1.01, -0.60]
large vs. control	-0.07	-0.39	.92	-0.04	[-0.24, 0.16]
small vs. control	1.45	7.64	< .001	0.77	[0.56, 0.97]
Ski					
large vs. small	-2.08	-7.20	< .001	-0.73	[-0.93, -0.52]
large vs. control	0.01	0.02	1.00	0.00	[-0.19, 0.20]
small vs. control	2.08	7.24	< .001	0.73	[0.52, 0.93]
Car					
large vs. small	-1.74	-6.04	< .001	-0.61	[-0.81, -0.41]
large vs. control	-1.23	-4.30	.001	-0.43	[-0.63, -0.23]
small vs. control	0.51	1.78	.82	0.18	[-0.02, 0.38]
Frequent flyer					
large vs. small	-0.25	-0.86	1.00	-0.09	[-0.28, 0.11]
large vs. control	-0.15	-0.51	1.00	-0.05	[-0.24, 0.15]
small vs. control	0.10	0.36	1.00	0.04	[-0.16, 0.23]
Fitness center					
large vs. small	-2.02	-6.98	< .001	-0.70	[-0.91, -0.50]
large vs. control	1.08	3.77	.01	0.38	[0.18, 0.58]
small vs. control	3.10	10.77	< .001	1.08	[0.87, 1.29]

Note. The *p* values of pairwise comparisons were adjusted using the Tukey's honestly significant difference method.

Cross scenario analyses: Combined mixed-effects findings for all scenarios

Tykocinski et al. (1995) reported an additional one-way ANOVA that combined the three scenarios in Experiment 2 (i.e., car, frequent flyer, and fitness center). The initial pre-registration planned to imitate this one-way ANOVA by collapsing all four scenarios together. However, upon further analysis we realized one-way ANOVA that does not account for the non-independence of the observations is not the best choice for this analysis. Also, because of the fact that the same participant might be in the large-difference condition in one scenario, but in the small-difference condition in another scenario, a mixed ANOVA would not fit well either.

In the end, we decided to go with a mixed-effects regression analysis (Bates, Marchler, Bolker, & Walker, 2015) which would properly handle the interdependent nature of the data. Psychologists have also recommended this method for analysis with repeated measures for which participants are assigned to different materials in different trials (Brysbaert & Stevens, 2018).

In particular, we created three dummy variables to represent the three conditions: the large-difference condition (Dummy 1), the small-difference condition (Dummy 2), and the control condition (Dummy 3). In a mixed-effects regression, we regressed the likelihood to act on the subsequent opportunity on the first two dummy variables, and used the third dummy as the reference category for comparison. We also included the random-intercept effects of scenario and subject.

Because there is no immediately obvious *p* value in mixed-effects regression, [1] following Bates et al. (2015), we estimated the statistical significance in mixed-effects regression through a logic of model comparison. That is, the difference in the chi-squared between two models with and without certain variables will serve as the effect size of these respective variables; the statistical significance of the change in chi-squared will indicate the statistical significance of the effect of these variables. To ease interpretation and comparison with previous analyses, we further converted chi-squared statistics to *F* statistics, Cohen's *f*, partial eta-squared, and Cohen's *d*. *F* statistics were calculated by dividing chi-squared by the degree of freedom. The other effect size measures were calculated based on Lakens (2013) and DeCoster (2012). We also obtained the confidence intervals of these statistics by mapping them with the noncentral chi-squared confidence interval.[2] The Cohen's *f*, partial eta-squared, and Cohen's *d* should be interpreted with caution because the conversions are based on a series of statistical assumptions that may not apply to mixed-effects regression.

---

[1] In order to handle unbalanced designs with multiple nested or fully crossed or partially crossed grouping factors for the random effects, mixed-effects regression is estimated by maximum likelihood (or residual maximum likelihood) rather than observed and expected mean squares as in ordinary linear regression or simple ANOVA (Bates, 2006).

[2] Exp1: The degree of freedom for pairwise comparison is estimated as 43/3*2-1=27.67. Exp2: The degree of freedom for pairwise comparison is estimated as 309/3*2-1=205.

Table K1

Results of Across Scenarios Analyses Using Mixed-Effects Regression

	$\chi 2$	*df*	Noncentral χ^2 95% CI - LL	Noncentral χ^2 95% CI - UL	*F*	*p*	*f*	η_p^2	η_p^2 90% CI - LL	η_p^2 90% CI - UL
Experiment 1										
Overall	11.48	2	1.39	27.34	5.74	.003	.54	.22	.05	.38
								Cohen's *d*	Cohen's *d* 95% CI - LL	Cohen's *d* 95% CI - UL
Large vs. Small	11.37	1	2.00	28.44	11.37	.001		-1.28	-2.03	-.54
Large vs. Control	4.07	1	.00 [a]	15.83	4.07	.04		-.77	-1.51	- [b]
Small vs. Control	2.03	1	.00 [a]	11.45	2.03	.15		.54	- [b]	1.29
Experiment 2										
	$\chi 2$	*df*	Noncentral $\chi 2$ 95% CI - LL	Noncentral $\chi 2$ 95% CI - UL	*F*	*p*	*f*	η_p^2	η_p^2 90% CI - LL	η_p^2 90% CI - UL
Overall	69.59	2	39.86	105.03	34.80	.00	.48	.19	.07	.14
								Cohen's *d*	Cohen's *d* 95% CI - LL	Cohen's *d* 95% CI - UL
Large vs. Small	69.46	1	40.63	105.97	69.46	.00		-1.16	-1.44	-.89
Large vs. Control	15.44	1	3.88	34.69	15.44	.00		-.55	-.82	-.28
Small vs. Control	20.06	1	6.34	41.45	20.06	.00		.63	.35	.90
Experiment 3 (Different Conditions Design)										
	$\chi 2$	*df*	Noncentral $\chi 2$ 95% CI - LL	Noncentral $\chi 2$ 95% CI - UL	*F*	*p*	*f*	η_p^2	η_p^2 90% CI - LL	η_p^2 90% CI - UL
Overall	58.41	2	31.43	91.08	29.20	.00	.31	.09	.06	.12
								Cohen's *d*	Cohen's *d* 95% CI - LL	Cohen's *d* 95% CI - UL
Large vs. Small	55.02	1	29.78	87.94	55.02	.00		-.86	-1.08	-.63
Large vs. Control	4.57	1	.00 [a]	16.79	4.57	.03		-.25	-.47	- [b]
Small vs. Control	28.20	1	11.22	52.85	28.20	.00		.61	.39	.84

Note. a. The lower limit has reached the smallest possible value, which is zero. b. Estimation is not available.

Table K2

Summary of Effect Comparisons Between the Target Article and the Replication Studies

Scenario	Statistical test	Target Article (*n* = 120)	Replication Experiment 1 (*n* = 43)	Replication Experiment 2 (*n* = 309)	Replication Experiment 3: Different Conditions (*n* = 603)	Conclusion
Across Scenarios	One-way ANOVA	*f* = .43	*f* = .54	*f* = .48	*f* = .31	Experiment 1 • Replicated: one-way ANOVA; large vs. small comparison. Experiment 2 • Replicated: one-way ANOVA; large vs. small comparison; large vs. control condition. • Not replicated: the small vs. control comparison was significant in the *opposite direction* of prediction. Experiment 3 (Different Conditions Design) • Replicated: one-way ANOVA; large vs. small comparison. • Not replicated: the small vs. control comparison was significant in the *opposite direction* of prediction.
	Large vs. Small	< 0	*d* = -1.28 [-2.03, -.54]	*d* = -1.16 [-1.44, -.89]	*d* = -.86 [-1.08, -.63]	
	Large vs. Control	< 0	*d* = -.77 [-1.51, NA]	*d* = -.55 [-.82, -.28]	*d* =-.25 [-.47, NA]	
	Small vs. Control	n.s.	*d* = .54 [NA, 1.29]	*d* = .63 [.35, .90]	*d* = .61 [.39, .84]	

Note. f = ANOVA effect size *f*; *d* = Cohen's *d* effects for two samples. n.s. = non-significant. The target article only provided information about the signs and significance levels of pairwise comparisons, but not the standard deviations/t-statistics/p values, thus we could not calculate the exact *d*s of the target article. Significance level is defined by $p < .05$, two-tailed test.

Experiment 1

The beta coefficient of the large vs. control comparison was -.98 (*S.E.* = .49, 95% CI = [-1.94, -.03]), and the beta-coefficient of the small vs. control comparison was .69 (*S.E.* = .48, 95% CI = [-.26, 1.63]). In a separate model where we included Dummy 1 and Dummy 3 (i.e., contrasting the large-difference condition and the control condition with the small-difference condition), the beta-coefficient of the large vs. small comparison was -1.67 (*S.E.* = .49, 95% CI = [-2.63, -.71]).

The overall effect of experimental conditions was significant (χ^2 (2) = 11.48, F = 5.74, f = .54, η_p^2 = .22, p = .003). The large vs. small comparison was negative and significant (χ^2 (1) = 11.37, d = -1.28, p < .001). The large vs. control comparison was negative and significant (χ^2 (1) = 4.07, d = -.77, p = .04). The small vs. control comparison was non-significant (χ^2 (1) = 2.03, d = .54, p = .15).

Experiment 2

The beta coefficient of the large vs. control comparison was -.77 (*S.E.* = .20, 95% CI = [-1.15, -.39]), and the beta-coefficient of the small vs. control comparison was .87 (*S.E.* = .19, 95% CI = [.49, 1.25]). In a separate model where we included Dummy 1 and Dummy 3 (i.e., contrasting the large-difference condition and the control condition with the small-difference condition), the beta-coefficient of the large vs. small comparison was -1.64 (*S.E.* = .19, 95% CI = [-2.02, -1.26]).

The overall effect of experimental conditions was significant (χ^2 (2) = 69.59, F = 34.80, f = .48, η_p^2 = .19, p = .000). The large vs. small comparison was negative and significantly (χ^2 (1) = 69.46, d = -1.16, p =.000). The large vs. control comparison was negative and significantly (χ^2 (1) = 15.44, d = -.55, p =.000). The small vs. control comparison was positive and significant χ^2 (1) = 20.06, d = .63, p =.000).

Experiment 3 (Different Conditions Design)

The beta coefficient of the large vs. control comparison was -.29 (*S.E.* = .14, 95% CI = [-.56, -.02]), and the beta-coefficient of the small vs. control comparison was .73 (*S.E.* = .14, 95% CI = [.46, .99]). In a separate model where we included Dummy 1 and Dummy 3 (i.e., contrasting the large-difference condition and the control condition with the small-difference condition), the beta-coefficient of the large vs. small comparison was -1.02 (*S.E.* = .14, 95% CI = [-1.29, -.75]).

The overall effect of experimental conditions was significant (χ^2 (2) = 58.41, F = 29.20, f = .31, η_p^2 = .09, p = .000). The large vs. small comparison was negative and significantly (χ^2 (1) = 55.02, d = -.86, p =.000). The large vs. control comparison was negative and significantly (χ^2 (1) = 4.57, d = -.25, p =.000). The small vs. control comparison was positive and significant χ^2 (1) = 28.20, d = .61, p =.000).

Summary

These suggest that across the four scenarios, there was a significant difference between the large-difference condition and the control condition, and between the large-difference condition and the small-difference condition. In Experiment 1 and Experiment 3 (Different Conditions Design) there was no support for effects between the small-difference condition and the control condition. In Experiment 2 there was support for effects between the small-difference condition and the control condition.

Results of One-way ANOVA for Experiment 3 (Same Condition Design)

For the purpose of comparison with results of one-way ANOVA for Experiment 1, Experiment 2, and Experiment 3 (Different Conditions Design), we also conducted one-way ANOVA for Experiment 3 (Same Condition Design). The results were largely consistent with those of mixed factorial ANOVA reported in the main text.

Table L1
Experiment 3 (Same Condition Design): One-way ANOVA on the Mean Likelihood of Acting on the Subsequent Opportunity

Experiment 3 Mixed factorial (N = 600)	Conditions									
Scenario	Large-difference	Small-difference	Control	*df*	*F*	*p*	*f*	η^2	90% CI for η^2	95% CI for η^2
Ski	5.09[a]	7.13[b]	5.08[a]	2, 597	32.66	< .001	.33	.099	[.06, .14]	[.06, .14]
Car	2.76[a]	4.49[b]	3.99[b]	2, 597	19.85	< .001	.26	.062	[.03, .09]	[.03, .10]
Frequent flyer	6.20	6.38	6.34	2, 597	0.23	.79	.03	.001	[.00, .01]	[.00, .01]
Fitness center	3.14[a]	5.13[b]	2.04[c]	2, 597	57.55	< .001	.44	.162	[.12, .20]	[.11, .21]

Note. For each scenario where the F test was significant, means of subgroups that do not share a common subscript are significantly different at $p < .05$. Pairwise comparisons were based on Games-Howell test.

Table L2

Experiment 3 (Same Condition Design): Pairwise comparisons of the likelihood of acting on subsequent opportunity following one-way ANOVA

Experiment 3 Mixed factorial (N = 600)	Mean Difference	*t*	*p*	Cohen's *d*	95% CI for Cohen's *d*
Ski					
large vs. small	-2.04	-7.31	< .001	-0.73	[-0.93, -0.53]
large vs. control	0.01	0.02	1.00	0.00	[-0.19, 0.20]
small vs. control	2.05	7.14	< .001	0.71	[0.51, 0.91]
Car					
large vs. small	-1.73	-6.30	< .001	-0.63	[-0.83, -0.43]
large vs. control	-1.23	-4.40	< .001	-0.44	[-0.64, -0.24]
small vs. control	0.50	1.71	.20	0.17	[-0.03, 0.37]
Frequent flyer					
large vs. small	-0.18	-0.67	.78	-0.07	[-0.26, 0.13]
large vs. control	-0.15	-0.50	.87	-0.05	[-0.25, 0.15]
small vs. control	0.32	0.11	.99	0.01	[-0.18, 0.21]
Fitness center					
large vs. small	-1.98	-6.47	< .001	-0.65	[-0.85, -0.45]
large vs. control	1.10	4.02	< .001	0.40	[0.20, 0.60]
small vs. control	3.09	10.48	< .001	1.05	[0.84, 1.25]

Note. Pairwise comparisons were based on Games-Howell test.

Mini Meta-Analysis using Restricted Maximum-Likelihood Estimator

We replicated the mini meta-analysis using a different approach based on mean and standard deviation using module MAJOR in JAMOVI. We again specified random effects models. Because there is essentially no one best estimator method to calculate between-study variance and weights (see Veroniki et al., 2016 for a review), we also chose a different estimator method, the restricted maximum likelihood method. Unlike the DerSimonian and Laird method which is non-iterative, the restricted maximum likelihood method is iterative. The results are presented in Figures M1 to M3.

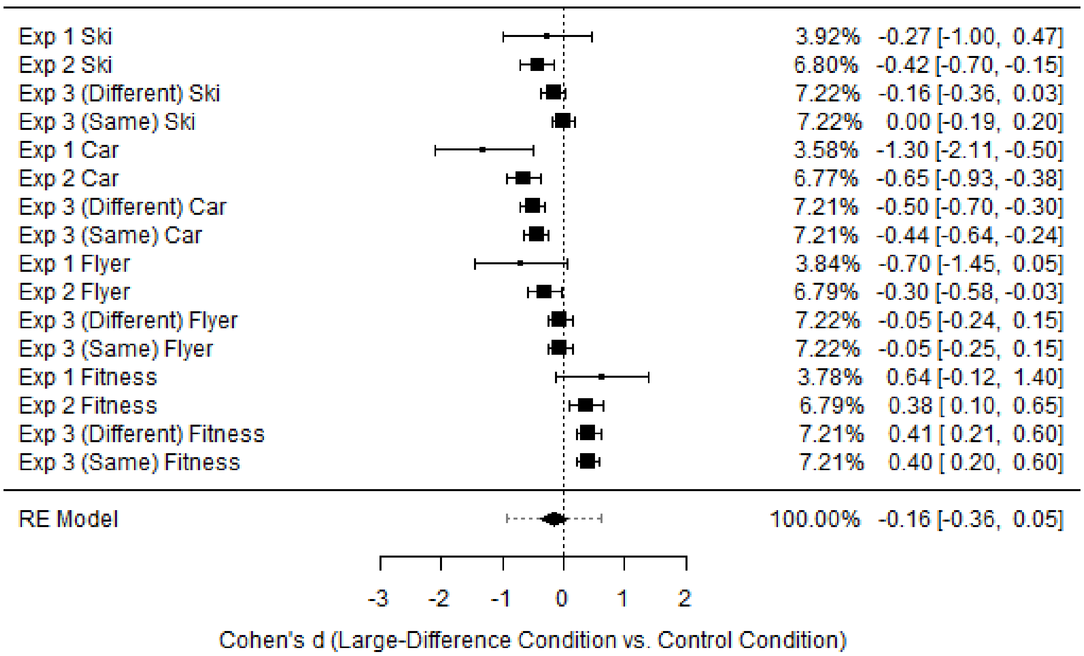


Figure M1. Forest plot: Large-difference condition vs. small-difference condition.

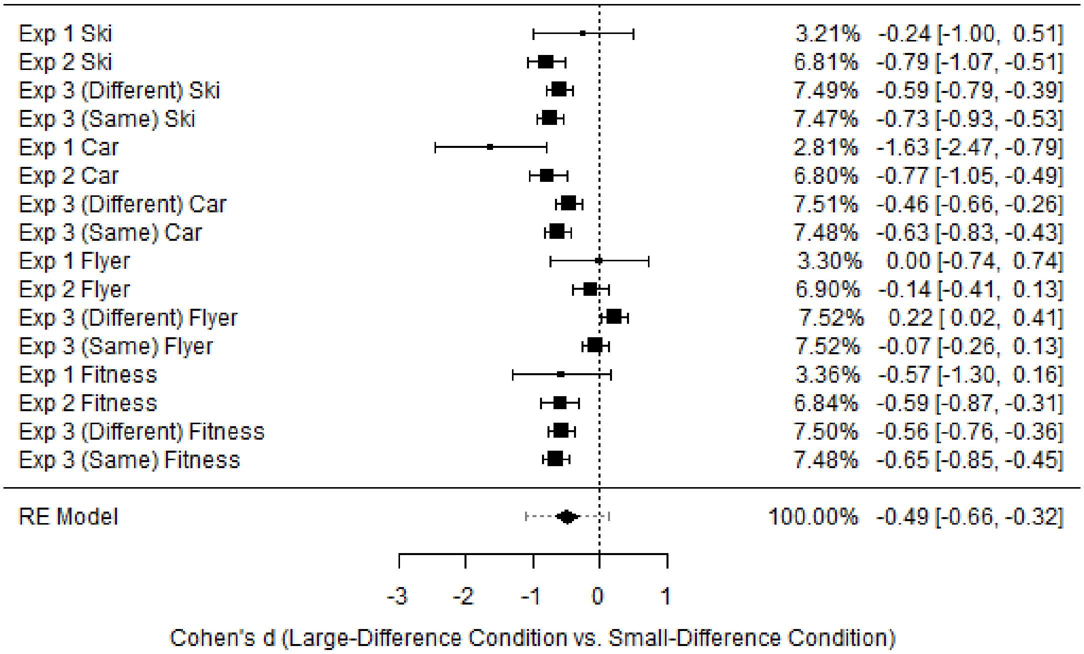


Figure M2. Forest plot: Large-difference condition vs. control condition.

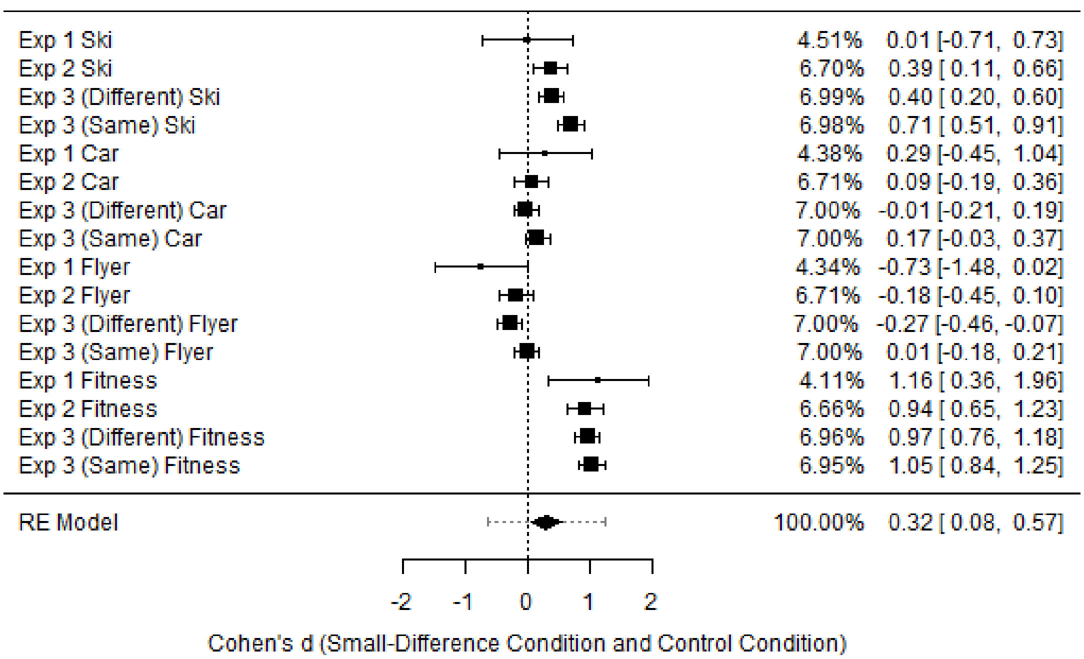


Figure M3. Forest plot: Small-difference condition vs. control condition.

Because the weights assigned to different samples in a random effects model vary when different sets of samples are put together (i.e., between-studies variance would differ, see Borenstein, Hedges, Higgins, Rothstein, 2009), we also conducted a separate set of mini meta-analysis for each individual scenario. The results are shown in Figures M4 to M6.

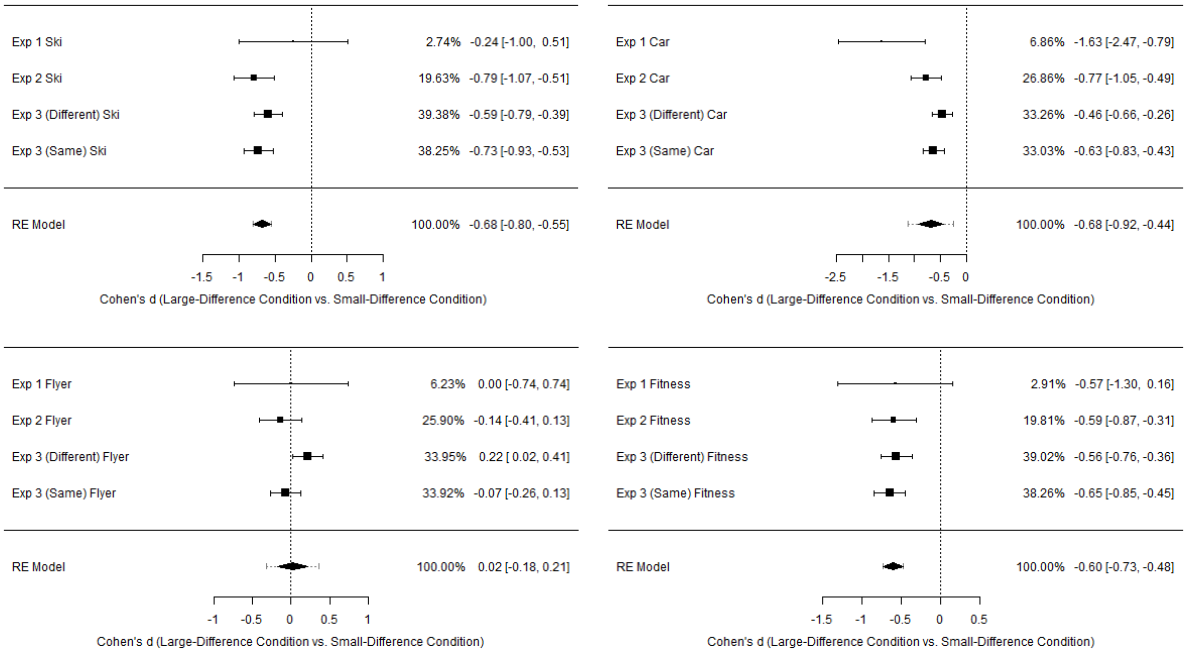


Figure M4. Forest plots for individual scenario: Large-difference condition vs. small-difference condition. From top to bottom, left to right, the forest plots are for ski, car, frequent flyer, and fitness center.

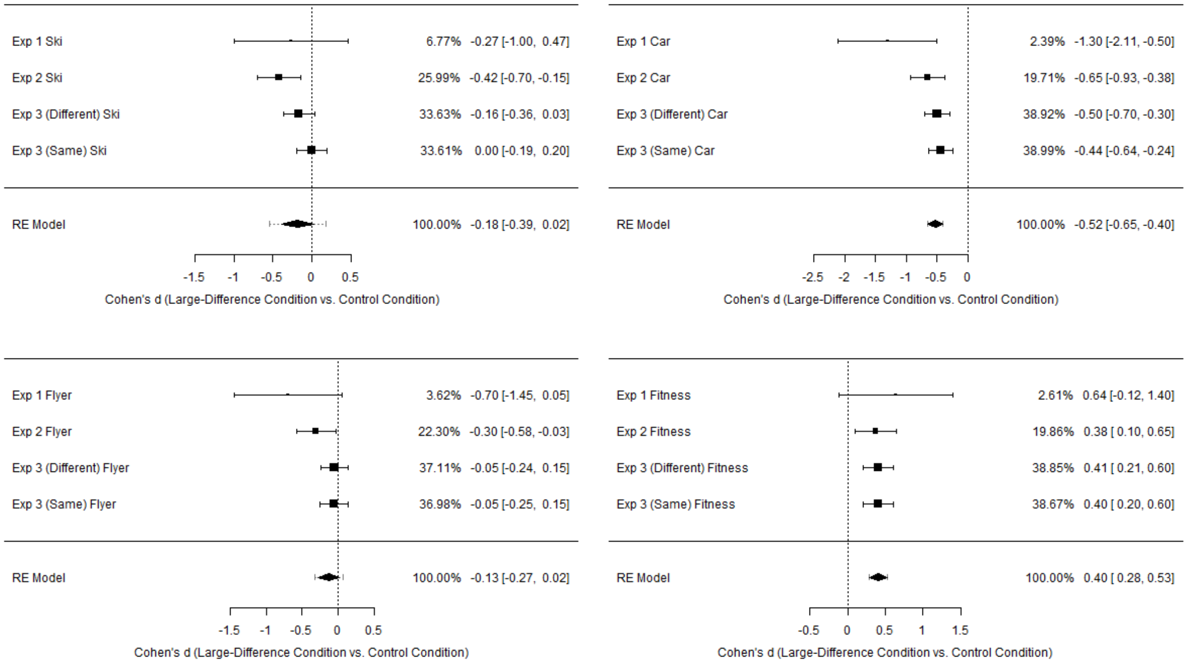


Figure M5. Forest plots for individual scenario: Large-difference condition vs. control condition. From top to bottom, left to right, the forest plots are for ski, car, frequent flyer, and fitness center.

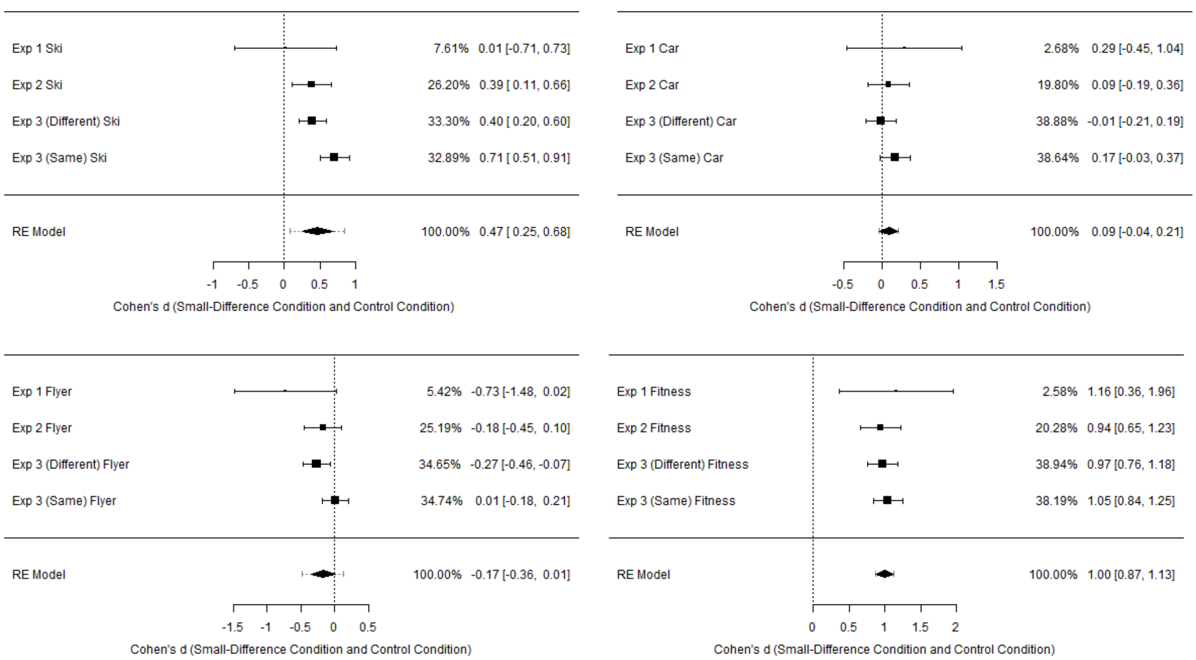


Figure M6. Forest plots for individual scenario: Small-difference condition vs. control condition. From top to bottom, left to right, the forest plots are for ski, car, frequent flyer, and fitness center.

Forest Plots for Ski, Car, Frequent Flyer Scenarios Only

Because the likelihood of acting in the control condition of the fitness center scenario was particularly low in all four samples, which stood in sharp contrast with the other three scenarios, we conducted a separate set of exploratory meta-analyses with data from ski, car, and frequent flyer scenarios only. We did not plan to run this set of meta-analyses before analyzing the data, and therefore urge readers to be cautious with making any conclusive interpretations from this set of analyses.

The results revealed significant difference between large-difference condition and small-difference condition (mini meta: *d* = -0.46, CI [-0.68, -0.23]) and between large-difference condition and control condition (*d* = -0.32, CI [-0.47, -0.17]), but not between small-difference condition and control condition (*d* = 0.11, CI [-0.08, 0.30]).

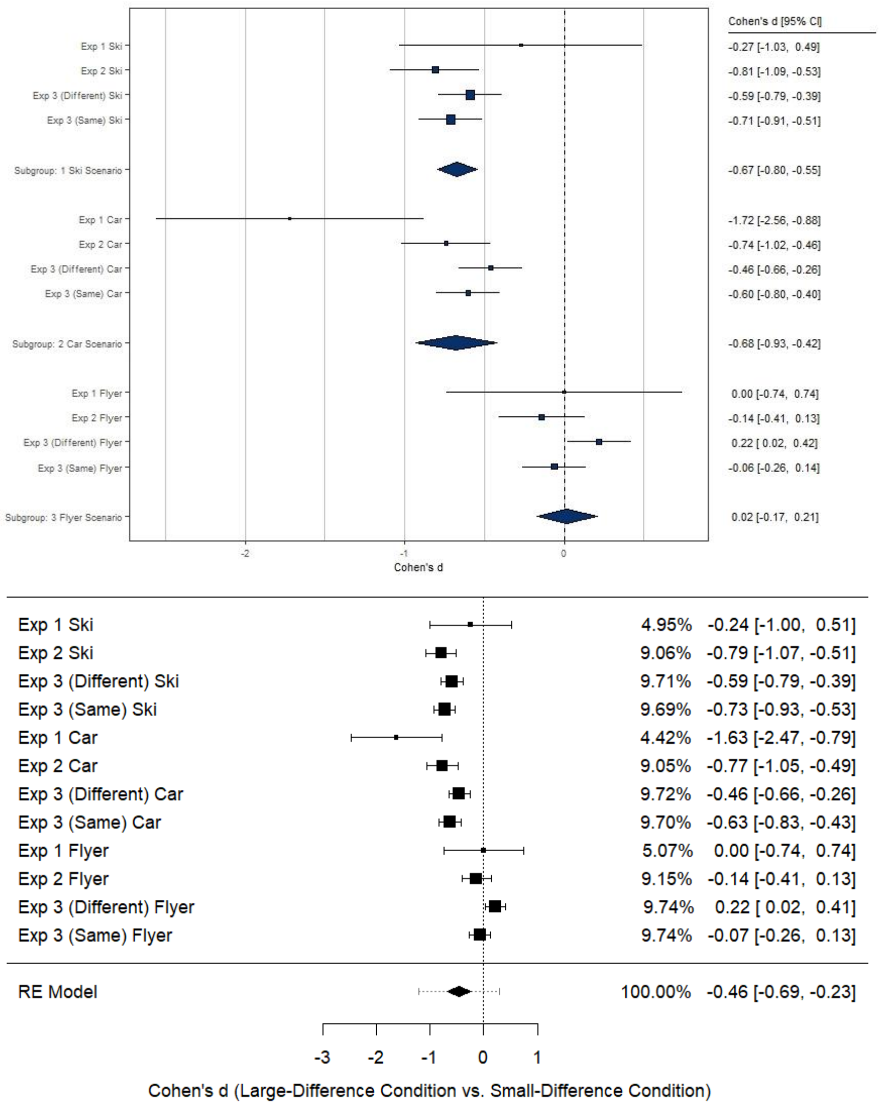


Figure N1. Forest plot: Large-difference condition vs. small-difference condition. Upper panel: meta-analysis based on effect sizes with DerSimonian and Laird estimator (overall effect size: *d* = -0.46, CI [-0.68, -0.23]); bottom panel: meta-analysis based on mean difference with restricted maximum-likelihood estimator.

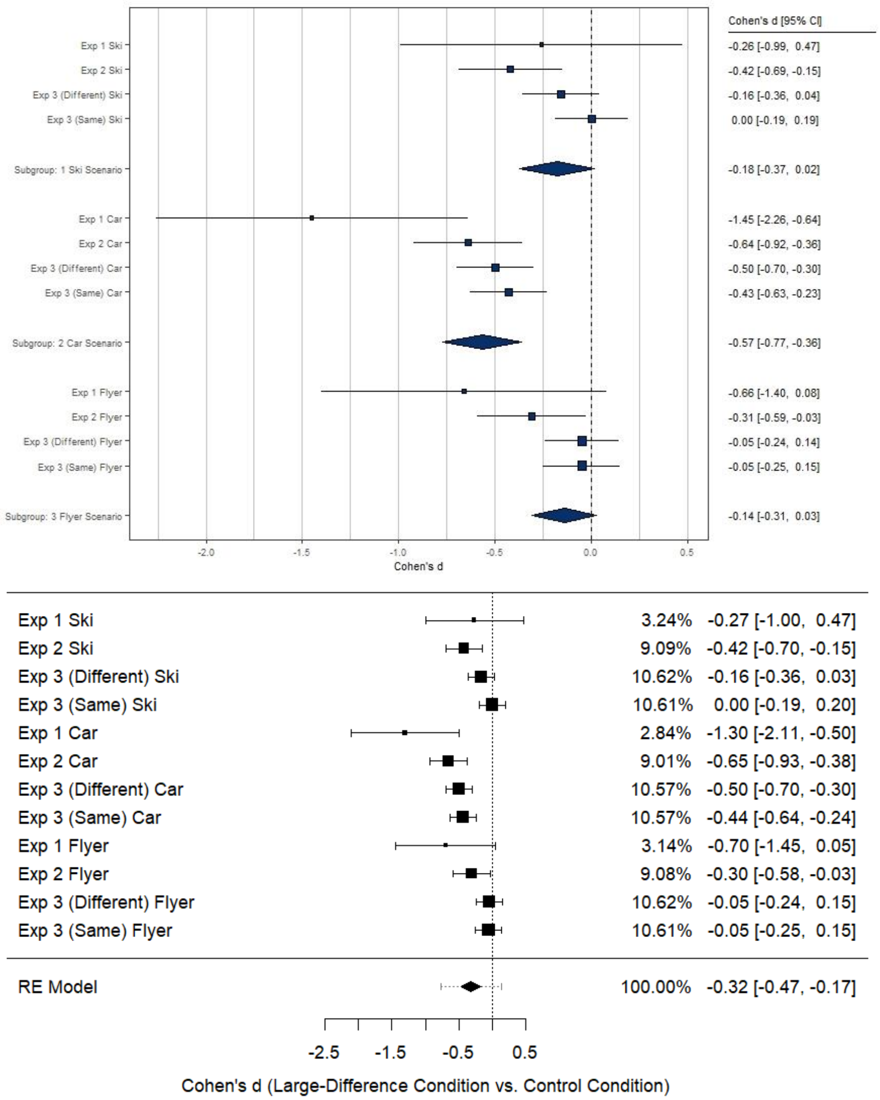


Figure N2. Forest plot: Large-difference condition vs. control condition. Upper panel: meta-analysis based on effect sizes with DerSimonian and Laird estimator (overall effect size: *d* = -0.32, CI [-0.47, -0.17]); bottom panel: meta-analysis based on mean difference with restricted maximum-likelihood estimator.

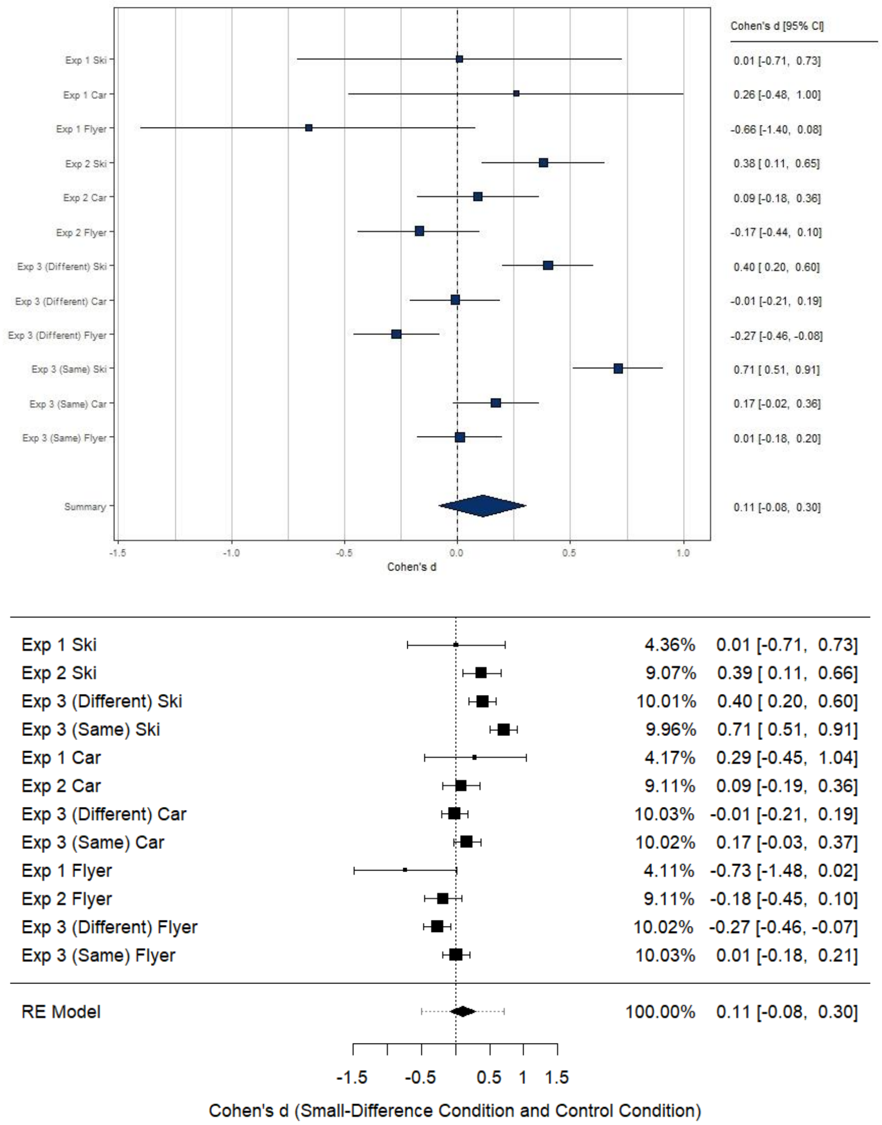


Figure N3. Forest plot: Small-difference condition vs. control condition. Upper panel: meta-analysis based on effect sizes with DerSimonian and Laird estimator (overall effect size: d = 0.11, CI [-0.08, 0.30]); bottom panel: meta-analysis based on mean difference with restricted maximum-likelihood estimator.

Notable Studies on the Inaction Inertia Effect

We conducted a review of the inaction inertia literature, and constructed the following summary table:

Table R1

A Comparison of Notable Studies on the Inaction Inertia Effect

Study	Same scenario?	All three conditions?	External lab?	Pre-registered?	Large vs. Control	Small vs. Control	*n* per condition
Tykocinski et al. (1995) Experiments 4-6	Yes	No	No	No			38~83
Tykocinski & Pittman (1998)	No	No	No	No			39-61
Tykocinski & Pittman (2001)	No	No	No	No			29-58
Tykocinski, Israel & Pittman (2004)	No	No	No	No			84
Pittman, Tykocinski, Sandman-Keinan, & Matthews (2008)	No	Yes	No	No	n.s.	n.s.	12~34
Terris & Tykocinski (2016)	No	Yes	No	No	Not reported	n.s.	40~62
van Putten, Zeelenberg & van Dijk (2007)	No	Yes	Yes	No	-	- or n.s.	21~40
Sevdalis, Harvey, & Yip (2006)	No	Yes	Yes	No	Not reported	Not reported	52~71
Krijnen, Zeelenberg,	No	No	Yes	No			90~227

Breugelmans, & van Putten (2019)							
van Putten, Zeelenberg & van Dijk (2009)	No	No	Yes	No			60~75
Zeelenberg, Nijstad, van Putten, & Dijk (2006)	Yes	Yes	Yes	No	-	-	21~80
Zeelenberg & van Putten (2005)	No	No	Yes	No			51~75

Note. n.s. = non-significant. When calculating the average sample size per condition, we excluded conditions that do not fit in the definition of inaction inertia effect, and took the averages of large-difference condition, small-difference condition, and/or control condition.