PNAS
www.pnas.org

Supplementary Information for:

The empirical audit and review and an assessment of evidentiary value in research on the psychological consequences of scarcity.

Authors: Michael O'Donnell[1*], Amelia S. Dev[2], Stephen Antonoplis[3], Stephen M. Baum[3], Arianna H. Benedetti[4], N. Derek Brown[3], Belinda Carrillo[3], Andrew Choi[3], Paul Connor[3], Kristin Donnelly[3], Monica E. Ellwood-Lowe[3], Ruthe Foushee[5], Rachel Jansen[3], Shoshana N. Jarvis[3], Ryan Lundell-Creagh[3], Joseph M. Ocampo[3], Gold N. Okafor[3], Zahra Rahmani Azad[6], Michael Rosenblum[7], Derek Schatz[8], Daniel H. Stein[3], Yilu Wang[9], Don A. Moore[3], Leif D. Nelson[3]

Affiliations:
[1] Georgetown University
[2] University of Miami
[3] University of California, Berkeley
[4] Airbnb
[5] University of Chicago
[6] University of Tuebingen
[7] New York University Stern School of Business
[8] Accenture, PLC
[9] Peking University

*Correspondence: michael.odonnell@georgetown.edu.

This PDF file includes:

> Supplementary text
> SI References

Supporting Information for: The empirical audit and review and an assessment of evidentiary value in research on the psychological consequences of scarcity.

Overall Method

Topic Selection

We carried out this collaborative research project at the University of California, Berkeley. The project leaders invited collaborators to propose research fields or findings worthy of a broad-based replication project. The project leaders screened suggestions to identify topic areas that would have a clear operational definition, would include a diverse set of investigations, and could be replicated within a reasonable budget and timeline. The reduced set of three possibilities (1. a selection of most cited papers within a specified time frame, 2. moral licensing, 3. the psychological effects of scarcity) was then presented to the research team, and by a vote, the third option was selected. The project leaders solicited example papers for the topic, and based on feedback from the team, refined the focus to replication of experimental manipulations of scarcity in particular. From there we collectively set down the exact parameters for the investigation.

We decided to use Shah, Mullainathan, and Shafir (2012) as the "seed paper" for identifying experimental studies of the psychological effects of scarcity. Our reasoning was both conceptual and practical. On the conceptual side, the publication of Shah et al. (2012) has been a watershed event in the study of scarcity. Although related papers preceded it, Shah et al. (2012) was perhaps the most prominent to both make claims that the psychology of scarcity can explain some aspects of the perseverance of poverty. Moreover, that paper was the first to suggest that some of those effects can be captured

with situational manipulations that activate the psychology of scarcity in populations that are otherwise not impoverished.

The selection was also practical. That paper has been very widely cited, and one could reasonably infer that any subsequent paper on the psychological effects of scarcity would be very likely to cite it. For that reason, we could start with a pool of papers that had cited Shah et al. (2012) and from that population of possible papers, winnow the selection down to those which met the exact specifications of our project. In essence, by starting with Shah et al. (2012) we could more readily land on a set of studies that were more likely to be centrally important and a selection strategy that could more easily be replicated. Our initial consideration set of papers was the list of all papers in the Web of Science Core Collection that cited the seed paper by January 28, 2019. This criterion yielded a set of 198 papers in the consideration set for replication.

Study Selection

Many of the 198 possible papers did not report any new experiments on the psychological effects of scarcity. Some were review papers, some were on an adjacent topic, and some were unrelated. We coded the full set of papers to identify any reported experiment that could meet the inclusion criteria for our investigation. Two collaborators were randomly assigned to each paper in the set and asked to independently determine its suitability for replication. Reviewers rated their assigned papers on three criteria: 1. are there experiments in the paper? 2. do the experiments have a scarcity manipulation, and if so, which experiments do? 3. can the studies that have a scarcity manipulation be replicated online (i.e., MTurk, the platform that we planned to use for data collection)? Reviewers

submitted their reviews to the first author, who determined agreement between the two assigned reviewers for each section. Most papers did not contain any experiments or were deemed unsuitable for replication online (166 / 198, 83.8%). Of the remainder, each was coded as either a "green light" paper, for which both reviewers agreed that the paper had appropriate studies and which studies met all three criteria, and "yellow light" papers for which only one reviewer indicated that one or more studies met all three criteria within the paper. This analysis yielded a selection of (23) green light papers and (9) yellow light papers. The OSF page for this project contains the original list of 198 papers as well as the reduced set of green or yellow light papers identified for potential inclusion. The first author verified which studies, if any, satisfied all three criteria for the "yellow light" papers in case these were necessary for use in assigning studies for replication.

With this pool of potential studies, we proceeded to randomly assign papers to each of the collaborators on the project. The 23 green light papers were each assigned a number, and then numbers 1 to 23 were drawn out of an urn, assigning a paper to each collaborator. In an effort to include as broad a swathe of scarcity research as possible, we did this initial randomization by paper rather than by study. Accordingly, for those papers with multiple eligible studies, we conducted a second drawing to determine which study in the paper we would replicate.

Each collaborator then closely reviewed the assigned study to determine whether the procedure could be faithfully replicated. If the collaborator determined that the assigned study was impractical for replication (e.g., because it required extensive use of a simulation the collaborators could not access), then a new paper was randomly selected,

and a random study identified. The final sample included 20 studies drawn from 20 different papers.

Pre-Registration and Analysis Plans

After we assigned studies for replication, each collaborator was also assigned a partner (another member of the collaborative team) who would be their "verifier." The verifier would work with the assigned replicator to ensure the fidelity of the replication to the original study. The verifiers and replicators worked together throughout the process. Each replicator designed their study to be a faithful replication of the original. The primary replicator was responsible for directing the replication of their study, while verifier was responsible for ensuring that the replication was faithful to the original study. Replicators and verifiers were encouraged to reach out to original authors for exact materials when the published reports had insufficient detail. After the replicator designed and programmed the study, verifiers compared study materials against the original, and offered feedback for correction of differences.

Before collecting any data, each replicator filed a pre-registration (on AsPredicted.org, and available on the OSF.io page for the project). All pre-registrations included the hypothesis and analysis plan from the original study, as well as the intended sample size for the replication. For all replications, we set a target sample size of 2.5 times that used in the original (Simonsohn, 2015). Verifiers checked the pre-registration against the original as well and co-authored the final preregistration.

Data Collection

Prior to beginning data collection, the first author verified the fidelity of each replication survey. Because we were running so many studies in parallel, it was impossible to completely exclude every participant from participating in more than one experiment. We came as close as possible to that goal. The first author made a subjective assessment of the commonality of the scarcity manipulations in each paper and identified three categories of manipulations that were used repeatedly, as well as a number of unique manipulations, and ensured that no participant was exposed to more than one experiment within each category of repeated manipulations. The three categories of repeat manipulations were: 1. individual financial hardship primes (5 studies), 2. economic recession primes (2 studies), and 3. scarcity primes for a product or resource (3 studies). For each of these three categories the studies were posted to MTurk in batches such that a participant could only complete one study in each batch. The remaining studies used manipulations that did not fall into any of these three broad categories and were posted individually.

Analyzing Results

Each replicator analyzed the data in accordance with their pre-registered analysis plan. Each of those initial analyses was further verified by an independent research collaborator (AD), who created the final posted analysis code available on the final OSF page for the project. We included and applied attention checks where they were used in the original studies. Although there has been some concern recently about bot activity on MTurk, we obviated this concern by including a captcha question on all studies run on the Qualtrics platform (except Layous, et al (2018), which was a longitudinal study requiring multiple weeks of participation). This included every study except two: the

replication of Camerer's replication of Shah et al. (2012) and the replication of Shah et al. (2015), which were run using proprietary scripts provided by the authors. However, bot activity does not represent the only potential threat to participant quality. As a check against careless responding, we manually flagged written responses for each of the six studies that used a writing task (Chou et al., 2016; Mani, Mullainathan, Shafir, & Zhao 2013; Mehta & Zhu, 2016; Plantinga, Krijnen, Zeelenberg, & Breugelmans, 2019; Roux, Goldsmith, & Bonezzi, 2015; Tully, Hershfield & Meyvis, 2015) as part of the independent variable if they appeared to be low quality. We used a light touch here and flagged responses that were either blank, nonsensical, or transparently copied and pasted from another source. For these six studies we report the results with and without these responses in the Supporting Information and the results without these responses in Figure 1 of the main text. All materials, data, and analysis code are available online: https://osf.io/a2e96/

Standardizing effect sizes

We sought to convert all key effect sizes, in both the original and replication study, to a standardized measure. This has the benefit of allowing us to draw meaningful comparisons between original and replication effect sizes. Following in the footsteps of large-scale replicability projects by Camerer et al. (2018) and the Open Science Collaboration (2015), we chose the correlation coefficient, *r*, as our standardized measure of effect size. As the Open Science Collaboration (OSC) writes: "Correlation coefficients are bounded, well known, and therefore more readily interpretable. Most important for our purposes, analysis of correlation coefficients is straightforward because, after applying the Fisher transformation, their standard error is only a function of sample size".

In most cases, we were able to convert effect sizes into correlation coefficients using R code provided by the Open Science Collaboration (OSC) from their 2015 replication project. You can view the function in an Appendix to the OSC project (https://osf.io/z7aux/). View the code we wrote to convert effects for our selected original studies and replication studies to correlation coefficients as an R Markdown file here: https://osf.io/5zrgf/. We calculated 95% confidence intervals around the correlation coefficients using the *CIr* function from the R package 'psychometric' (Fletcher, 2010). The knitted Markdown file can be downloaded and opened in your browser as an HTML file (https://osf.io/7zhnd/). For convenience, see below for the formulas used in the OSC's function:

The z statistic is transformed into a correlation using sample size N:

$$r_f = \tanh\left(z\sqrt{\frac{1}{(N-3)}}\right)$$

with r_f being the Fisher-transformed correlation. The χ^2 statistic is transformed into the phi coefficient with:

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

The t and F statistic are transformed into a "correlation per df" using:

$$r = \sqrt{\frac{F\frac{df_1}{df_2}}{F\frac{df_1}{df_2}+1}\sqrt{\frac{1}{df_1}}}$$

where $F = t^2$. The expression in the first square root equals the proportion of variance explained by the df_1 predictors of the variance not yet explained by these same predictors. To take into account that more predictors can explain more variance, we divided this number by df_1 to obtain the "explained variance by predictor". Taking the square root gives the correlation, or more precisely, it gives the correlation of each predictor assuming that all df_1 predictors contribute equally to the explained variance of the dependent variable.

The OSC function can be used for calculating correlation coefficients from analyses reporting $z, F, t,$ or χ^2 statistics. For original or replication studies that did not report results of a $z, F, t,$ or χ^2 test, we converted existing results into one of the above statistics first. These studies included Study 4 of Lee et al. (2018), Experiment 2 of Mehta & Zhu (2016), and Study 3 of Plantinga et al. (2018). See below for details on those cases. All correlation coefficients can be found in columns I (original) and M (replication) of our Scarcity Replication Project spreadsheet. The correlation coefficient for the replication study was coded as negative if the effect was in the opposite direction from the original study.

Small telescopes analysis

The small telescopes analysis (Simonsohn, 2015) allows us to estimate the power of any given original study using the replication effect size. It also allows us to ask what the required sample size would be to obtain 80% power for detecting the replication effect size. We conducted this analysis in R, within the same Markdown file containing the code to calculate correlation coefficients. More specifically, for each study, we performed the following steps for the small telescopes:

1. Used the R function *pwr.r.test* from the package 'pwr' (*14*) to estimate power of the original study given the original study sample size and the replication study effect size (*r*)

2. Used *pwr.r.test* to estimate the sample size needed for 80% power to detect the replication effect size (*r*) at a significance level of .05

3. Confirmed the power results of *pwr.r.test* from step 1 by simulating data with the original study sample size and replication effect size and estimating power from the simulations. Estimated power was calculated as the proportion of iterated simulations that returned a significant result. For estimating power, we ran 50,000 iterations of each simulation.

4. Confirmed the sample size results of *pwr.r.test* from step 2 by simulating data, with the replication effect size, testing for the power obtained at different sample sizes. This allowed us to loosely confirm the *pwr.r.test* results for the sample size needed to obtain 80% power for detecting the replication effect.

5. Repeated steps 1 – 4 for the upper bound of the 95% CI around the replication effect sizes.

The simulations were conducted using the *grid_search* function from Jeffrey Hughes' R package 'paramtest' (*16*), with guidance from Hughes' accompanying vignette here (*16*) and from a blog post on using simulations to estimate power in R, here (*17*).

Unplanned deviations and analysis exceptions

Replication study authors selected a key hypothesis from each original study, preregistered this hypothesis as the hypothesis of interest in the replication, and preregistered the same statistical test the original authors used to test their hypothesis. However, there were some unplanned deviations from this protocol. In the Bickel et al. (2016), Lee et al. (2018), and Mehta & Zhu (2016) projects, alternative key hypotheses,

meant to isolate the effect of the scarcity manipulation (and thus alternative hypothesis tests) were identified, representing unplanned deviations from the replication preregistrations and requiring an attempt to derive parallel results from the original study. Errors in the replication survey materials also led us to repeat the Bickel et al. and Emery et al. replications. The Camerer et al. (2018) replication had a unique sample size deviations that we detail below. Finally, although we did not preregister a particular method of deriving correlation coefficients (r), we think it is worth providing additional detail on how we computed r in cases where the OSC function could not be used.

Bickel et al. (2016)

Data collection deviations.

After running the first replication, we realized we were uncertain if our calculation of ln(k) was correct. We wrote the original authors and asked for their help understanding the results of our first replication. The original authors shared relevant materials, which we used to correct our survey and the calculation of ln(k) for a second replication attempt. All results we present in the manuscript are from our second replication attempt, conducted with the corrected survey materials. More detail on the ln(k) calculation error, as well as the findings from our first (flawed) replication attempt, can be found in the Bickel et al. (2016) project write-up (https://osf.io/ecvfa/).

Analysis deviations.

The critical result is reported in Table 3 of the original paper (Bickel et al., 2016). This table separately reports model results for each of the four trial types (gain future, gain past, loss future, loss past). However, we are interested only in the overall effect of

scarcity and collapsed across all four trial types. Additionally, although the original paper reports the results of negative, neutral, and positive income shock conditions, we chose to focus on the difference between only the negative and neutral income shock conditions. Thus, for our key hypothesis test, we tested for an effect of income shock condition (negative vs. neutral) on delay discounting, aggregating data across all trial types (future gain, future loss, past gain, past loss).

The original authors publicly shared their data, meaning we were able to run this reduced model on our replication study data and on the original study data. All results reported in the manuscript for the Bickel project are the results of this reduced model. Effect sizes were calculated using the F-value for the scenario (income shock condition) variable in these models. Full results for the reduced model are also included in the Bickel write-up linked above.

Camerer et al. (2018) replication of Shah et al. (2012) Study 1

Data collection deviations.

Camerer et al. (2018) is, itself, a replication project. We replicated the Camerer et al. (2018) replication of Shah et al. (2012). Camerer et al. (2018) established that they would run one replication and then, should that replication fail to yield significant results, collect additional data in order to reach 90% power to detect 50% of the effect they sought to replicate. Camerer et al. (2018) did not find a significant effect in their first replication sample of 278 participants, and hence, recruited an additional sample of 341 for a full sample size of 619. We based the power analysis for *our* replication on Camerer et al.'s first sample (N = 298), yielding our replication sample size of 668 (after

13

exclusions). However, we realized after our replication was complete that Camerer et al. established their final effect estimate based on their full sample of 619 participants. Thus, the "original result" we compare to in our manuscript is the statistical result from Camerer et al.'s (2018) analysis in their full sample.

Chou et al. (2016)

Data collection deviations.

We preregistered a sample size of 653 and ended data collection with 659 responses. After manually reviewing and removing 45 careless responses (e.g., responses that were completely nonsensical or did not attempt to answer the prompt), 603 valid responses remained. Results both with and without exclusions for careless responding are reported below.

Emery et al. (2015) Study 3

Data collection deviations.

We ran this replication study twice. However, we were unable to analyze results from the first sample (N = 380) due to the pre-registered exclusion of participants who were currently in a romantic relationship. In the first sample, 292 participants (~75% of the recruited sample) indicated they were in a romantic relationship. With the addition of the other two pre-registered exclusions (see the preregistration for the first replication attempt here), the first sample included only N = 47 viable participants. This sample size did not allow us to adequately assess replicability. Therefore, we sought to conduct a second replication. In this collection, participants were pre-screened to ensure that only

participants who were not in a relationship at the time of participation were allowed to complete the study. Data collected from the initial replication was not used in final analyses. After data collection for the second replication attempt, and again following preregistered exclusion criteria (see the preregistration for the second replication attempt here), the final sample included N = 339 participants.

Layous et al. (2018)

Data collection deviations.

In correspondence with the original study authors, we learned that the original study (Layous, Kurtz, Chancellor, & Lyubomirsky, 2018) was conducted on a website custom-programmed by one of the study authors. The website was no longer available at the time of our replication; however, the original study authors shared a PDF of their materials, which we used to implement the study in Qualtrics.

Analysis deviations.

We contacted the original study authors for help understanding how they implemented their primary hypothesis test, a multilevel linear model. The original authors shared the SPSS code and R code they used in their data cleaning, scoring, and analysis. From their code, it appears that for participants who did not have complete data for the well-being composite at all three time points (baseline, post-intervention, and 2 weeks post-intervention), well-being score was coded as "system missing" when running the multilevel linear model. Additionally, the well-being composite was only computed for participants with a certain threshold of complete data for each scale used in the composite, representing another level of potential missing data. This is not reported in the

original study paper, nor is the final sample size after accounting for these "system missing" data-points.

As authors were not comfortable sharing their data, we were unable to calculate the exact sample size after "system missing" exclusions. To the best of our ability, we ascertained a sample size of N = 111 after exclusions, which is the sample size the original authors report after accounting for participants who failed to complete the third and final time point. Of note is that we are unclear if N = 111 includes or excludes participants who did not complete the second time point. There may be participants who completed time point three but did not complete time point two, for example, meaning our estimate of N = 111 for the multilevel linear model could be inflated, as only participants with complete time point data were included in the model. Additionally, even within the subset of participants who completed all three time points, there may have been further exclusions for missing data on one or more of the scales used to compute the well-being composite. Thus, while we use N = 111 for our effect size estimation, that number should be taken as our best good faith estimate rather than as definitive.

For the replication study, we conducted an identical version of the linear mixed effects model mentioned above, using a replication dataset that also excluded participants who did not complete well-being measures at all 3 time points. The results reported in the manuscript therefore represent results after excluding participants with incomplete data for any of the three time points, for both the replication and, to the best of our knowledge, for the original study.

Lee et al. (2018) Study 4

Analysis deviations.

Primary hypothesis test. The preregistered analysis was a 3 x 2 ANOVA testing for main effects of consumption resources (unearned vs. earned vs. unspecified), fairness (high vs. low), and their interaction on brand attitudes. Upon review of the study details, we decided that in order to test for effects of scarcity we would focus on the attenuated interaction of the reduced model: consumption resources (earned vs. unearned) x fairness (high vs. low) on brand attitudes. In the replication, we were able to implement this reduced model by simply sub-setting our data to exclude participants in the unspecified resources condition. The replication effect reported in the manuscript is the effect of the resources x fairness interaction in this reduced dataset. However, we needed a way to derive this result from the available information in the original study paper.

Calculating the test statistic for the original study. There was enough information to estimate the *F*-value from the reduced 2 x 2 ANOVA in Table 3 of the original study (Lee, Baumgartner, & Winterich, 2018). To do this, we implemented the protocol from pages 195 – 197 in Cohen's 2002 paper (*15*), "Calculating a Factorial ANOVA From Means and Standard Deviations". Calculations were done in Excel, and you can view the Excel file with the steps and results here. We recommend referencing the Cohen paper and the linked Excel file for more information, but below, we also explain this estimation in terms of the Lee study.

The original study reported a sample size of 251 eligible participants, in a six-cell design. The original paper does not report participants per cell, meaning we had to estimate the per cell sample size at 251 ÷ 6, or 41.83. Skewing conservative, we rounded

down to 41 participants per cell for an estimated total N of 164 (41 x 4 cells) in our reduced 2 x 2 model.

We first calculated marginal means (row and column means) from the cell means reported in Table 3 of the original study. Using the cell standard deviations reported in Table 3 and our marginal means to Cohen's Formula 5 (2002):

$$SS = N\sigma^2$$

we calculated between-cell, row, and column sums of squares. In this formula, N is the total sample size and σ^2 represents the biased variance of the respective means. So, for example, in calculating row sums of squares, the formula was applied as follows:

$$SS_{rows} = N(\sigma_{rows})^2$$

This calculation was repeated for the between-cells sums of squares ($SS_{between-cells}$) and column sums of squares (SS_{cols}), using their respective biased variances. Following Cohen (2002), we then calculated sums of squares for the interaction as:

$$SS_{interaction} = SS_{between-cells} - SS_{rows} - SS_{cols}$$

F-ratio numerators can be calculated by dividing each sum of squares ($SS_{interaction}$, SS_{rows} and SS_{cols}) by its respective degrees of freedom. In the case of the Lee study, this was always 1, so the *F*-ratio numerators were the pure sums of squares values.

As noted above, because we did not know the true per-cell N we assumed a balanced design. Thus, the *F*-ratio denominators were calculated as the average of all cell variances. From there, *F*-values for the rows (fairness), columns (consumption resources), and the interaction were calculated by dividing the numerator and denominator. The Excel F.DIST.RT function was used to estimate an associated *p* value for each *F* statistic.

To confirm the results derived from the formulas above, we generated a simulated sample, with identical summary statistics as those from Table 3 (Lee et al., 2017), using the *rnorm* function in R. We then ran the same 2 x 2 ANOVA on the simulated data. The *F*-statistic for the interaction effect in the simulated data was $F = 2.652$, identical to the *F*-statistic for the interaction calculated as calculated in Excel using Cohen's formulas. View the R code here: https://osf.io/xmbwt/.

Calculating effect size. The *F*-value for the interaction, being our key statistic of interest for the original study, was converted to a correlation coefficient using the OSC function. Finally, partial eta-squared for the main effect of consumption resources, the main effect of fairness, and the effect of their interaction, was calculated as:

$$\eta_\rho{}^2 = \frac{SS_{between-cells}}{SS_{between-cells} + SS_{within-cells}}$$

which can be converted to this formula:

$$\eta_\rho{}^2 = \frac{Fdf_{between-cells}}{(Fdf_{between-cells)} + df_{within-cells}}$$

where within-group degrees of freedom ($df_{within-cells}$) was calculated as N (sample size) – k (number of groups): 164 – 4 = 160. And, as mentioned, $df_{between-cells}$ was equal to 1.

Mani, Mullainathan, Shafir, & Zhao (2013) Study 4

Data collection deviations.

We preregistered a sample size of 240 and ended data collection with 236 responses. After manually reviewing and removing 7 careless responses (e.g., responses that were completely nonsensical or did not attempt to answer the prompt), 229 valid responses remained. Results both with and without exclusions for careless responding are reported below.

Mehta & Zhu (2016) Experiment 2

Data collection deviations.

We preregistered a sample size of 383 and ended data collection with 387 responses. After manually reviewing and removing 8 careless responses (e.g., responses that were completely nonsensical or did not attempt to answer the prompt), 381 valid responses remained. Results both with and without exclusions for careless responding are reported below.

Analysis deviations.

Primary hypothesis test. The preregistered replication analysis was a binary logistic regression testing for the effect of scarcity condition (resource scarcity, resource abundance, or control) on the correctness of solutions on the Duncker Candle task. Although not explicitly preregistered, we believe that the main test of interest is the comparison of just the scarcity and control conditions. The original study (Experiment 2 of Mehta & Zhu, 2016) reports results comparing just the scarcity and control conditions, meaning we were able to deriving our effect size of interest from the original study results relatively simply. And, of course, for the replication study, we were able to run the condition comparison on a subset of our data that included only the scarcity and control conditions.

Calculating effect size. The test of interest in the original study, a binary logistic regression, was reported with the Wald statistic as an effect size measure. We could not find any definitive source on how to convert Wald statistics, specifically, to correlation coefficients. As Wald tests have a chi-squared distribution under the null hypothesis, we assumed that the Wald statistic would be treated as a chi-squared statistic for the sake of converting the effect size. Thus, in applying the OSC function to convert correlation coefficients for the Mehta & Zhu (2016) original and replication study, we applied the chi-squared formula to the Wald statistics associated with condition in the binary logistic regressions.

The original study reported overall sample size (N = 133) but not sample size per condition. As we were interested in only the scarcity vs. control conditions, we estimated N per condition as $133 \div 3 = 44$ (rounded down from 44.3$\underline{3}$). Thus, in converting the

effect size for the original study, we estimated the total sample size as N = 88 (44 x 2 conditions).

For the replication, we were again able to subset our data to exclude the abundance condition and include only the scarcity and control condition. The sample size, after removing those in the abundance condition, was N = 236, which is the value we used in estimating the correlation coefficient.

Plantinga et al. (2018) Study 3

Data collection deviations.

We preregistered a sample size of 1605 but ended up recruiting 1613 participants from MTurk. Manual review identified 59 careless responses (e.g, nonsensical responses or responses that did not attempt to address the study prompt), resulting in a sample size of 1564 valid responses. Results both with and without exclusions for careless responding are reported below.

Analysis deviations.

Calculating effect size. The test of interest in the original study (Study 3 in Plantinga, Krijnen, Zeelenberg, & Breugelmans, 2019) was a logistic regression examining the effect of condition (opportunity cost reminder or control), centered effective income, and their interaction on buying decisions. The result of interest is the odds ratio for the effect of the income x condition interaction, which can be found in Figure 2 of the original paper. We converted the odds ratio (*OR*) for the original and

replication study results to an approximated point-biserial correlation (r) using the following formula (Bonett, 2007):

$$\frac{ln\ ln\ (OR)}{\left\{ln\ ln\ (OR) + \frac{2.89n^2}{n_1 n_2}\right\}^{\frac{1}{2}}}$$

Plantinga et al. uploaded their data on OSF (https://osf.io/gks5q/), meaning we were able to replicate their Study 3 logistic regression with their data. This allowed us to use the exact, un-rounded *OR* for our original study effect size conversion. Conversions from *OR* to *r* were conducted in a separate R script (https://osf.io/m7j2p/). As well as converting *OR* to *r*, the R script includes a formula for converting *OR* to Cohen's *d*, a formula for calculating 95% CIs around *d*, and a formula for converting *r* to *d* as a means of double checking our results (Borenstein et al., 2009).

Roux, Goldsmith, & Bonezzi (2015) Study 4

Data collection deviations.

We preregistered a sample size of 393 and ended data collection with 422 responses. After manually reviewing and removing 8 careless responses (e.g., responses that were completely nonsensical or did not attempt to answer the prompt), 414 valid responses remained. Results both with and without exclusions for careless responding are reported below.

Shah et al. (2018) Study 3

Analysis deviations.

The replication preregistration specified that participants should be excluded for missing click data; however, the replication authors did not exclude these participants when originally running their analyses. We repeated analyses on the replication study after these exclusions. The results in the manuscript represent results from the replication analyses *with* exclusions and are reflected in the updated project write-up (https://osf.io/6r7gc/). Older results can be found in the project archive (https://osf.io/nmk8r/).

Tully et al. (2015) Study 5

Data collection deviations.

We preregistered a sample size of 1,018 and ended data collection with 1,012 responses. 300 participants were removed based on the pre-registered exclusion criteria. After manually reviewing and removing 24 careless responses (e.g., responses that were completely nonsensical or did not attempt to answer the prompt), 688 valid responses remained. Results both with and without exclusions for careless responding are reported below.

Analysis deviations.

The replication preregistration did not specify whether the repeated measures ANOVA should be conducted with the assumption of compound symmetry in the covariance structure. However, we believe that the key hypothesis test in the original study (Study 5 of Tully, Hershfield, Meyvis, 2015) was conducted with the assumption of

compound symmetry, which is a default setting in some statistical packages. The initial

replication analyses were conducted using the *aov* function in R, which does not assume

compound symmetry. As such, we repeated analyses using the R *aov_car* function

(Singmann, 2010), which does assume compound symmetry (code here:

https://osf.io/psd7y/). The results in the manuscript represent this updated replication

analysis. The updated project write-up is here: https://osf.io/k6ytw/. The findings from

the first replication analysis, which did not assume compound symmetry, can be found in

the project archive here: https://osf.io/fdqta/.

Succinct Summaries of Methods and Results from Original and Replication Studies

Abeyta, Routledge, Kersten, & Cox (2017) Study 2

Original: Does making financial insecurity salient cause a reduced feeling of meaning in life? Participants were 121 undergraduates who were assigned to view a slideshow that ostensibly summarized a new article about either an economic recession or architectural building styles in this two-cell design. After viewing the slideshow, participants completed a version of the presence of meaning subscale from the Meaning in Life Questionnaire on a 9-point scale. Participants in the recession condition had a lower presence of meaning score than those in the control condition ($M = 5.46$, $SD = 1.68$ vs. $M = 6.04$, $SD = 1.47$; $F(1,119) = 3.98$, $p = .048$, eta2 = .03).

Replication: We recruited 304 workers to complete this study on Mechanical Turk. We randomly assigned participants to one of two conditions in a 2-cell (slideshow: economic recession (n = 158) or architecture (n = 146)) between-subjects design. We submitted these responses to a one-way ANOVA analysis, which revealed no significant differences in MLQ between the economic scarcity (M = 5.87, SD = 2.07) and control conditions (M = 6.09, SD = 2.01; F(1,302) = 0.91, $p = 0.34$.

Bickel, Wilson, Chen, Koffarnus, & Franck (2016)

Original: Does an economic narrative indicating a negative income shock increase the rate of temporal discounting in the past and future, for both gains and losses? Does a positive income shock narrative decrease the rate of temporal discounting in the past and future, for both gains and losses? Does making a zero explicit (e.g., $0 now and $100 in one month) interact with the effects of a positive or negative income shock? Participants

were 599 adults recruited through MTurk in this 3 (economic narrative: positive, negative, or neutral) X 2 (zero: explicit or implicit) design. Participants first envisioned themselves experiencing a negative income shock (losing their job and having no income), a positive income shock (a promotion with a 100% increase in income), or a neutral income shock (a lateral move with a 2% raise). Then, participants completed 4 sets of delay-discounting tasks which had participants consider gains and losses in the past and in the future. Briefly, participants were asked to choose between gaining or losing $500 now and $1,000 in the future or past. Participants in the explicit zero condition saw, e.g., $500 now and $0 in three months, while this in the implicit zero condition saw $500 and $0 in three months. The second time window ($1,000) began at 3 months and was titrated to go up or down systematically on the basis of participant's choices for 5 trials, producing a final discounting parameter, k, as described in further detail by Koffarnus & Bickel (2014). Results indicated that negative income shocks were associated with greater delay discounting than both neutral and positive income shocks, but that neutral and positive income shocks did not differ from one another in their effects on participants' discount rate. Moreover, the implicit zero produced higher discount rates in all conditions.

Replication: We recruited 1,518 workers to complete this study on Mechanical Turk. Participants were randomly assigned to one of 6 conditions in this 3 (income narrative: negative, neutral, or positive) X 2 (zero: explicit or implicit) between-subjects design. Immediately after reading the income narrative scenario, participants were asked to rate their mood on a 7-point scale from 1 (very sad) to 7 (very happy). Following this, each participant completed the 5-trial adjusting delay discounting procedure for four

conditions (future gain, future loss, past gain, past loss) in randomized order, to obtain the discounting parameter, k. This parameter was then natural log-transformed for each condition separately, resulting in a measure of delay discounting, $\ln(k)$, for each participant for each condition. For our primary analyses, we tested the effects of a Type 2 ANOVA. These analyses were performed using the ANOVA function in the package car, version 2.1-6 (Fox & Weisberg, 2011). Eta-square effect size and confidence intervals were subsequently calculated using Type 3 ANOVA, with the function eta_sq in the package sjstats, version 0.14.3 (Lüdecke, 2018). We identified as the primary test of interest the effect of scenario on the discount rate. Our test revealed a significant effect of scenario on participants' discounting rate ($F(1,984) = 10.72$, $p = 0.001$, η2p = .002). The effect of employment on discounting rates was also significant ($F(1,984) = 12.61$, $p <$ .001, η2p = .022).There was no significant effect of zero framing ($F(1,999) = 0.54$, $p =$ 0.46, η2p = .000).

Camerer et al. (2018) replication of Shah et al. (2012) Study 1

Original: Does scarcity cause greater cognitive fatigue, measured by poorer performance on a cognitive ability task? Participants were 619 adults recruited from MTurk, assigned to one of four conditions in a 2 (scarcity: rich or poor) X 2 (borrowing: present or absent) fully between-subjects design. All participants completed a 14-round Wheel of Fortune (WoF) game. Participants assigned to the poor conditions were given 84 guesses (6 per round) to correctly guess letters in the WoF game while participants assigned to the rich conditions were given 280 guesses (20 per round). Participants in the borrowing conditions could borrow against a future round for additional guesses in the current round. After completing the WoF game, all participants completed a Dots-Mixed task,

which required fixating on a pattern to the left or right of a central point. Congruent trials required pressing a key on the same side as the pattern and incongruent trials required pressing a key on the opposite side of the pattern. An ANOVA predicting participants' responses to the Dots-Mixed task revealed that poor participants performed worse than rich participants, $F(1,617) = 0.1389$, $r = -0.015$, $p = .710$.

Replication: We recruited 733 workers to complete this study on Mechanical Turk. We excluded 25 participants for not fully completing the study and we excluded an additional 40 from analyses for having no correct responses on the Dots-Mixed task. We assigned participants to one of four conditions in a 2 (scarcity: rich or poor) X 2 (borrowing: present or absent) fully between-subjects design. All participants completed the 14-round Wheel of Fortune (WoF) game. After completing the WoF game, all participants completed a Dots-Mixed task, which required fixating on a pattern to the left or right of a central point. The mean completion time for all 668 participants included in analyses was 25 minutes (SD=13 minutes). After exclusions, there were 322 participants in the poor condition and 346 participants in the rich condition. An ANOVA predicting participants responses to the Dots-Mixed task revealed that poor participants ($M_{poor} = 41.56$, $SD_{poor} = 16.79$) performed no worse than rich participants ($M_{rich} = 42.82$, $SD_{rich} = 15.87$) $F(1,666) = .99$, p=.319.

Chou, Parmar, & Galinsky (2016) Study 3

Original: Does economic insecurity cause physical pain? Participants were 231 adults recruited from MTurk in this two-cell design. All participants were asked to complete a short autobiographical writing task. Participants assigned to the high insecurity condition

wrote about a time when their economic prospects were uncertain (i.e., were financially insecure), while those assigned to the low insecurity condition wrote about a time when they felt certain about their economic prospects (i.e., were financially secure). All participants then completed an adapted version of the McGill Pain Questionnaire, which assessed pain in the head, chest, and stomach on an 11-point scale. Participants also completed the Positive and Negative Affect Schedule as a control measure. A *t*-test revealed that participants who wrote about a time of high financial insecurity experience more pain than those who wrote about a time of low financial insecurity. ($M = 0.86$, $SD = 1.54$ vs. $M = 0.44$, $SD = 1.08$, respectively; $t(229) = 2.38$, $p = .01$, $d = 0.31$.)

Replication: We recruited 659 participants to complete this study on Mechanical Turk. A manual review of careless responding indicated that 45 participants submitted non-sensical replies to the writing prompt, leaving a sample size of 603 responses. Participants were assigned to one of two conditions in a 2-cell (insecurity: low insecurity or high insecurity) between-subjects design. We told participants that they would be engaging in several different surveys that were independent from one another. All participants were first asked to write about an autobiographical experience. Participants then responded to three items that measured how much pain they were experiencing in their head, chest, and stomach on an 11-point scale anchored at (0 = no pain and 10 = worst pain ever experienced). These items are adapted from a short form of the McGill Pain Questionnaire (Melzack, 1987). We averaged participants' responses to these three items to form an index of physical pain ($\alpha = .85$). Next, participants completed the PANAS. We averaged participants' responses on items that measured negative affect to form an index of negative affect ($\alpha = .94$). Finally, participants also provided their age,

gender, and current employment status. The results of an independent samples *t*-test indicated that participants who recalled an economically insecure period in their life reported more physical pain (M = 1.57, SD = 2.14) than participants who recalled an economically secure period (M = 0.78, SD = 1.49, $t(601) = 5.24$, $p < .001$, Cohen's $d = 0.43$, 95% confidence interval (CI) = [0.27, 0.59]. This effect did not remain significant after we controlled for participants' age, gender, current employment status, and negative affect, $p = .08$. Including careless responders (N = 45) did not change qualitative results: participants who recalled an economically insecure period in their life reported more physical pain (M = 1.94, SD = 2.57) than participants who recalled an economically secure period (M = 1.06, SD = 1.99, $t(656) = 4.89$, $p < .001$, Cohen's $d = 0.38$, 95% confidence interval (CI) = [0.26, 0.56]) and this effect did not remain significant after we controlled for participants' age, gender, current employment status, and negative affect, $p = .14$.

Cook and Sadeghein (2018) Study 3

Original: Does a "triple scarcity effect" (no liquidity, loss frame, and lack of other lending options) lead to increased intent to borrow from a payday lender to cover $500 in expenses? Participants were 199 American adults recruited from MTurk and assigned to one of 8 conditions in this 2 (liquidity: absent or present) X 2 (frame: loss or gain) X 2 (lending options: present or absent) fully between-subjects design. Liquidity was manipulated by telling participants they were either broke (absent) or were recently paid (present), frame was manipulated by telling participants they were either behind on their auto bills and risked repossession (loss) or wanted an upgrade (gain), and lending options were manipulated by telling participants a payday loan was their only source of cash

(absent) or they had other options but preferred a payday loan (present). All participants were told that they needed to borrow $500 to cover their expenses and were able to borrow up to $1,000 from a payday lender. An ANOVA predicting the amount participants borrowed from their condition assignment (all three factors plus interaction terms) returned a marginally significant three-way interaction, $(F(1, 189) = 3.00, p = .08,$ $eta2 = .016)$. Participants who were assigned to the triple scarcity condition requested to borrow an average of $725, an amount that was higher than any of the other conditions.

Replication: We recruited 507 workers to complete this study on Mechanical Turk. We randomly assigned participants to one of eight conditions in a 2(liquidity: present or absent) X 2(consequence frame: gain or loss) X 2 (other lending options: present or absent) in a fully between-subjects design. All participants read a prompt which varied according to condition assignment. For the liquidity conditions we told participants either, "you just got paid" (present) or "you are broke" (absent). For the consequence frame conditions we told participants either, "you want to lease a new car model to benefit from the additional features" (gain) or "you are four months behind on your car and now risk repossession" (loss). For the other lending options conditions we told participants "you have no other lending options, so a payday loan is your only source of cash" (present) vs. "you have other lending options, but you still decide to acquire cash through a payday loan" (absent). We placed these prompts immediately before the description of how a payday loan works, but after we collected demographic variables. All participants were told that they needed to borrow $500 to cover their expenses and were able to borrow up to $1,000 from a payday lender. Because participants were told in the prompt that they could only borrow up to $1,000, we excluded from analysis all

participants who requested to borrow more than $1000. We submitted participants responses to an ANOVA predicting the amount participants borrowed from their condition assignment (all three factors plus their interaction), which returned a significant three-way interaction, (F(1, 497) = 9.75, p = 0.002, partial2p= .019).

Durante, Griskevicius, Redden, & White (2015) Study 4

Original: Does scarcity lead families to make greater financial investments in girls than in boys? Because girls represent a surer reproductive bet (i.e., less risky) does trait risk aversion predict an even stronger tendency to preferentially invest in girls under poorer economic conditions? Participants were 198 adults recruited from MTurk and assigned to one of two conditions in this two-cell design. All participants viewed a slideshow about the economy; half the participants saw a slide show depicting an economic recession and half saw a slide show depicting an economic upswing. Participants completed both a three-item measure of their preference for leaving different types of resources (savings, real estate, and valuables) in their will to a girl or a boy, and a six-item measure of risk-aversion that required them to indicate their relative preference for a sure thing over a risky prospect on a 6-point scale. Participants also completed a number of other belief measures that tested hypotheses secondary to the main focus of this study. A repeated measures ANOVA with each of the three items on the resource questionnaire as a within-subjects factor, economic condition as a between-subjects factor, and a continuous measure of risk aversion returned a significant effect of economic condition, such that participants allocated more resources to girls than to boys (F(1, 192) = 5.54, P = 0.02, d = 0.34 and a significant interaction between economic condition and the individual measure of risk aversion (β = .29, F(1, 192) = 6.33, P = 0.01), indicating that this difference was

reduced for people who were less risk averse. [Info about the process model from the original paper: As an additional test of the hypotheses in this study, the authors used the Hayes (2008, model 1) PROCESS procedure to probe the interaction by examining the effect of economic condition at 1 SD above and below the mean of risk aversion (Aiken and West 1991). At high levels of risk aversion (1 SD above the mean), economic recessions (vs. economic upswings) led participants to increase the overall allocation of assets to girls to a far greater degree (Mrecession= 4.83 vs. Mupswing= 4.35; t(196) = 3.44, P < 0.001). However, at low levels of risk aversion (1 SD below the mean), there was no effect of economic condition on bequeathing overall assets to a girl versus a boy (Mrecession = 4.56 vs. Mupswing = 4.58; t(196) = .12, P > 0.90).]

Replication: We recruited 573 workers from Mechanical Turk to complete this study of the effect of economic condition (recession vs. upswing) on resource allocation to daughters at the expense of sons. We randomly assigned participants to one of two conditions in this 2-cell (economic condition: recession or upswing) between-subjects design. All participants viewed a slideshow (we contacted the authors and used the same slideshow as the original study) displaying either an economic recession or an economic upswing and were asked to imagine themselves in that economy. Participants then imagined they were allocating assets in a will and indicated their preference to give their savings, real estate, and other valuables to their son versus their daughter and then indicated their preference to six gamble-like choices as a measure of risk aversion. We conducted a repeated measures ANOVA with asset type as a within-subjects factor, economic condition as a between-subjects factor, and a continuous measure of individual differences in risk aversion. There was no significant effect of economic condition, $F(1,$

530) = 0.78, P = 0.377. There was a significant main effect of risk aversion with participants higher in risk aversion more likely to allocate to sons at the expense of daughters, $F(1, 530) = 32.96$, P < 0.001. The key hypothesis test of the authors' theory, the interaction between economic condition and risk aversion, was significant, $\beta = 0.08$, $F(1, 530) = 3.59$, P = 0.059, 95% CI [0.00, 0.16]1. To test the directionality of the hypothesis, following the analyses of the original paper, we used model 1 from PROCESS (Hayes, 2017). Participants lower in risk aversion were more likely to allocate resources to daughters in the economic recession condition as compared to the economic upswing condition, b = -0.19, $t(530) = -2.00$, P = 0.046, 95% CI [-0.37, 0.00], and there was no significant difference between conditions for participants higher in risk aversion, b = 0.11, $t(530) = 1.01$, P = 0.312, 95% CI [-0.10, 0.32]. While the replication did find a significant interaction, it was not in the same direction and the original effect size was not included in the replication 95% confidence interval.

Emery, Walsh, & Slotter (2015) Study 3

Original: Does being primed with reduced self-concept clarity (SCC) cause less spontaneous self-expansion? Further, does reduced SCC attenuate an otherwise positive relationship between motivation to affiliate and self-expansion? Participants were 152 heterosexual, single adults recruited through MTurk in a three-cell design. All participants first completed a self-rating task in which they indicated how much 10 neutral attributes described them. Participants were then assigned to either generate two inconsistent self-aspects and write about how they sometimes conflict with each other in the participants' day-to-day life (*SCC threat*), or to generate two consistent self-aspects

and write about how they complement each other in the P's day-to-day life (*SCC confirmation*). Participants in the control condition wrote about their trip to get to the study. Then, under the cover of developing an online dating service, participants viewed a profile of an opposite sex target individual that listed several characteristics of the individual, including an attribute that the participant indicated *did not* describe them at the beginning of the study. Finally, participants indicated their interest in meeting the individual, and re-answered the self-rating task. Self-expansion was operationalized as an increased endorsement of the attribute the participant indicated *did not* describe them but was characteristic of the target in the dating profile. This difference was accounted for in a residualized analysis, controlling for initial ratings. An ANCOVA predicting participants' responses to re-rating of the target attribute controlling for their initial rating, found that participants in the SCC threat condition rated this attribute as less characteristic of them than did participants in the SCC confirmation or control conditions, $F(2, 148) = 39.24$, $p < .001$, eta2 $= .21$. There was no difference between the confirmation or control conditions. Moreover, a regression analysis with balanced contrast codes found a significant interaction effect between SCC condition and desire to meet the target individual in the dating profile on SCC expansion, beta $= .18$, $t(147) = 2.13$, $p < .05$. In the SCC confirmation and control conditions, a greater desire to meet the individual was associated with greater self-expansion, but there was no such relationship in the threat condition.

Replication: We recruited 339 workers to complete this study on MTurk. Participants were assigned to one of three conditions in this 3-cell (self-control clarity: threat, confirmation, or control) design. Participants completed the writing task and then

completed a measure of Self-Concept Clarity, Self-Esteem, and Affect. Then, under the

cover of developing an online dating service, participants viewed a profile of an opposite

sex target individual that listed several characteristics of the individual, including an

attribute that the participant indicated *did not* describe them at the beginning of the study.

Finally, participants indicated their interest in meeting the individual, and re-answered

the self-rating task. Self-expansion was operationalized as an increased endorsement of

the attribute the participant indicated *did not* describe them, but was characteristic of the

target in the dating profile.

We first checked whether the manipulation altered participants' perceptions of having

clear and cohesive identities. Contrary to the original paper, a between-subjects ANOVA

revealed a non-significant effect of assigned condition on SCC, $F(2, 331) = 0.562$, $p =$

.571, partial $\eta2 = .003$. Planned contrasts showed no differences in reported SCC in the

threat condition compared to the confirmation, $t(331) = 0.145$, $p = .885$, or control, $t(331)$

$= -0.848$, $p = .397$, conditions. Likewise, participants did not report differences in SCC in

the confirmation condition compared to participants in the control condition, $t(331) = -$

0.980, $p = .328$. Thus, we failed to replicate the basic manipulation from the original

paper. The SCC manipulation did not influence participants' self-esteem, $F(2, 331) =$

1.377, $p = .254$, partial $\eta2 = .008$. The manipulation also did not influence participants'

positive mood, $F(2, 329) = 0.501$, $p = .606$, partial $\eta2 = .003$, negative mood, $F(2, 329) =$

1.336, $p = .264$, partial $\eta2 = .008$, or interest in meeting the target individual, $F(2, 329) =$

1.319, $p = .269$, partial $\eta2 = .008$. Similar to the original paper the SCC manipulation did

not operate through influencing self-esteem, mood, or romantic approach motivations.

We next examined the primary hypothesis that the manipulation influenced participants'

spontaneous self-expansion. A between-subjects analysis of covariance revealed a non-significant effect of condition on self-ratings of the target attribute participants generated at the beginning of the study after viewing the dating profile controlling for initial self-ratings, $F(2, 138) = 0.9377$, $p = .394$, partial $\eta2 = 0.0134$. Planned contrasts revealed no significant differences in the endorsement of the target attribute between participants in the threat condition compared to those in the confirmation, $t(138) = 0.638$, $p = .524$, or control, $t(138) = -0.049$, $p = .961$, conditions. Likewise, participants in the confirmation condition did not endorse the target attribute more than participants in the control condition, $t(138) = -1.149$, $p = .252$.

Fernbach, Kan, & Lynch (2015) Study 2

Original: Do participants' planned responses to an unexpected constraint vary based on the magnitude of the constraint? Specifically, does a greater constraint lead participants to shift towards a planning strategy focused on priority planning (characterized by a focus on opportunity cost) relative to efficiency planning (characterized by a focus on stretching resources), and does a greater constraint also increase the speed of generating priority plans relative to efficiency plans? Participants were 102 adults recruited from MTurk and assigned to one of three conditions in this three-cell design. All participants were recruited during the holiday season and asked to list five people whom they would give a gift to and to write down a gift idea, price, and provide a URL for where they could buy each gift. Participants were next randomly assigned to one of three conditions in which they were told they received an unexpected bill in the mail for either $100, $500, or $1,000. Participants then wrote down how their plan for each of the gift recipients would change, as well as any other changes they would make to their shopping

plan. Independent coders categorized each plan as either an efficiency plan (19%),

prioritization plan (27%), or various other types of behavior (54%). An ANOVA

predicting the number of plans from plan type (efficiency vs. priority) and a linear

contrast of increasing financial burden ($F(1, 99) = 7.54$, $p = .007$) showed a planning mix

shift toward priority planning with increased constraint.

Replication: We recruited 255 workers from Mechanical Turk to complete this study. We

assigned participants to one of 3 conditions in this 3-cell (expense: $100, $500, or

$1,000) between-subjects design. All participants were asked to list five people whom

they would give a gift to and to write down a gift idea, price, and provide a URL for

where they could buy each gift. Participants were next randomly assigned to one of three

conditions in which they were told they received an unexpected bill in the mail for either

$100, $500, or $1,000. Participants wrote down how their plan for each of the gift

recipients would change, as well as any other changes they would make to their shopping

plan. Responses were coded as efficiency plans (such as finding sales or deals) or priority

plans (roughly involving trade-offs or sacrifices), or other types of behavior. The authors

predicted increased constraint would shift the planning mix toward prioritization and

would increase the speed of generating priority plans relative to efficiency plans. We

submitted participants' responses to an ANOVA predicting planning mix from constraint

and plan type. The ANOVA did not return a significant interaction of constraint and plan

type on planning mix, $F(1,222) = 0.29$, $p = .752$. In testing for response times, there was

an interaction of constraint and plan time on log-transformed response time, $F(2, 299) =$

3.04, $p = .049$. Following the original authors, we also ran a random intercept model "to

explore the relative accessibility of priority and efficiency plans" as specified in

Appendix E of Fernbach, Kan, & Lynch (2015). In testing for response times, there was no interaction of constraint and plantype on log-transformed response time, ($\beta = -.17$, $z = -1.61$, $p = .109$). This is contrary to the authors' results for an identical model, which found a significant interaction between constraint and plan type, ($\beta = -.13$, $z = -2.03$, $p = .042$).

Kristofferson, McFerran, Morales, & Dahl, (2017) Study 5

Original: Do scarcity promotions activate aggressive tendencies in consumers? Is the relationship between scarcity promotions and aggressive tendencies moderated by perceived social threat? Participants were 194 adults recruited from MTurk who were randomly assigned to one of four conditions in a 2 (promotional ad: scarcity vs. control) X 2 (perceived competitive threat: high vs. low) fully between-subjects design. First, all participants completed the threat manipulations. Participants were asked to either write down two ways in which they were similar to other consumers in their city (low threat) or different from other consumers in their city (high threat). Next, participants were exposed to the scarcity manipulation. All participants saw an advertisement for an iPhone 5 that was offered at a reduced price and were told there were only 3 (scarcity) or 3,000+ (control) available at their local electronics store. Then, all participants read about lining up to participate in the sale and were told that they arrived before the store opened and were positioned ahead of other consumers near the front of the line. Finally, in an ostensibly unrelated study, participants selected a video game they wanted to play from among each of 7 pairs of Super Nintendo video games. Each pair featured one violent video game and one non-violent game. The dependent variable of aggression was conceptualized as the proportion of violent video games participants selected to play in

their seven choices. A two-way ANOVA with proportion of violent video games as the dependent variable and contrast coded values for each of the manipulated factors entered as predictors returned a significant main effect of perceived threat, $F(1, 156) = 4.14$, $p = .044$ and an interaction effect between perceived threat and scarcity promotion, $F(1, 156) = 3.96$, $p = .048$.

Replication: We recruited 502 workers from MTurk to complete this study. Participants were randomly assigned to one of four conditions in a 2 (promotional ad: scarcity, control) x 2 (perceived competitive threat: high, low) between-subjects design. We modified the original materials upon consultation with the original authors. The original experiment was conducted when the iPhone 5 was available for sale. We conducted the replication in April 2019 when the iPhone 5 was no longer a "new" promotional item. We utilized the same exact promotional ad as in the original study, except we used the iPhone X. Participants were asked to either write down two ways in which they were similar to other consumers in their city (low threat) or different from other consumers in their city (high threat). Next, participants were exposed to the scarcity manipulation. All participants saw an advertisement for an iPhone X that was offered at a reduced price and were told there were only 3 (scarcity) or 3,000+ (control) available at their local electronics store. Then, all participants read about lining up to participate in the sale and were told that they arrived before the store opened and were positioned ahead of other consumers near the front of the line. Finally, in an ostensibly unrelated study, participants selected a video game they wanted to play from among each of 7 pairs of Super Nintendo video games. Each pair featured one violent video game and one non-violent game. The dependent variable of aggression was conceptualized as the proportion of violent video

games participants selected to play in their seven choices. We excluded responses from 51 participants who skipped through the scenario without reading (operationalized as spending less than 2.5 seconds on the page) and 70 participants who failed the instructional attention check. Thus, the analysis is conducted using 389 participants who completed the study as designed. The pattern of results does not differ if all participants are used. The manipulation check was successful. Participants perceived the iPhone promotional package quantity to be more scarce in the scarcity versus the control ad (1 = very scarce to 7 = very abundant; MScarcity = 4.44 vs. MControl = 1.80; $F(1, 385) = 249.20$, $p < .001$, partial eta-squared = .392). No main effect ($F(1, 385) = 2.31$, $p = .12$, partial eta-squared = .004) or interactions with perceived threat emerged ($F(1, 385) = 1.31$, $p = .253$, partial eta-squared = .003). We contrast-coded both promotion ad (–1 = scarcity, 1 = control) and perceived threat (–1 = low, 1 = high) independent variables and entered them in a two-way ANOVA with the proportion of violent games selected as the dependent variable. The ANOVA revealed no effect of perceived threat (PHigh Threat = .30, SD = .24 vs. PLow Threat = .34, SD = .26, $F(1, 385) = 2.65$, $p = .105$, partial eta-squared = .007) and no effect of scarcity promotion (PScarcity = .33, SD = .25 vs. PControl = .32, SD = .25, $F(1, 385) = 0.04$, $p = .843$, partial eta-squared = .000). In addition, the hypothesized interaction was not significant ($F(1, 385) = 0.01$, $p = .945$, partial eta-squared = .000). Participants in the scarcity–high-threat condition did not express a higher preference for violent games than participants in the control–high-threat condition (PScarcity-High Threat = .29 vs. PControl – High Threat = .30, $t(385) = 0.174$, $p = .862$, $d = 0.03$). In addition, no differences between scarcity and control ads emerged among participants in the low-threat condition (PScarcity – Low Threat = .34 vs.

PControl – Low Threat = .34, t(385) = 0.08, p = .933, d = 0.01). In addition, participants

in the scarcity – high-threat condition did not express a higher preference for violent

video games than participants in each of the other three conditions individually

(PScarcity – High Threat = .29 vs. PScarcity – Low Threat =.34, t(385) = -1.17, p = .242,

d = -.17; PScarcity – High Threat= .29 vs. PControl – High Threat = .30, t(385) = -0.17, p

= .862, d = -0.026; PScarcity – High Threat= .29 vs. PControl – Low Threat = .34, t(385)

= -1.29, p = .199, d = -.19) as well as against the average of other three conditions

(F(1,387) = 1.16, p = .283, partial eta-squared = .002). When participants who failed the

attention check were included, the predicted interaction remains consistent and non-

significant (F(1,498) = 0.68, p = .410, partial eta-squared = .001).

Layous, Kurtz, Chancellor, & Lyubomirsky (2018)

Original: Does imagining time as scarce increase wellbeing? Participants were 139

undergraduates who were assigned to one of two conditions in this two-cell design.

Participants who were assigned to the *live this month* (LTM) condition were told to live

the next month like it was their last in their college town for a while, and that they should

try to appreciate and savor the activities they do while they do them. Participants in the

control condition were told to keep track of what they do over the course of the week.

Participants completed an initial session, a weekly check-in over the course of four weeks

in which they wrote for 8 minutes about their activities during the prior week, and a

follow-up two weeks after the final session, for a total study duration of 6 weeks.

Participants completed the Brief Measure of Psychological Needs, the Satisfaction with

Life Scale, and the Modified Differential Emotions Scale; responses on the latter two

scales were combined into a well-being (WB) composite score. The researchers estimated

multilevel growth models to assess within-person changes in WB over time and condition differences in WB trajectories. Participants in the LTM condition showed steeper gains in linear wellbeing than the control group, $\gamma 11 = 0.11$, SE $= 0.06$, t(228) $= 2.00$, p $= 0.05$, 95% CI [0.002, 0.22], d $= 0.25$, which also showed linear gains in well-being over time, $\gamma 10=0.09$, SE $= 0.04$, t(228) $= 2.13$, p $= 0.03$, 95% CI [0.007, 0.16], d $= 0.21$. Thus, by the end of the intervention, participants in the LTM condition increased in well-being by nearly a half a standard deviation (d $= 0.46$), whereas those in the control group increased in well-being by less than half as much (d $= 0.21$).

Replication: We recruited 443 workers from MTurk to complete this study. We told participants that we were conducting a 6-week study that would require them to complete a survey that involved writing and answering questions once a week. Each time, they had to complete the survey within 24 hours in order to receive payment. We paid participants $1 for each survey, including the initial survey. Upon entering the study, participants completed two measures of well-being: The Satisfaction with Life Scale (SWLS) and the Modified Differential Emotions Scale (mDES). We then randomly assigned participants into one of two conditions in this two-cell design (Live This Month or Control). In the Live This Month (LTM) condition, participants were instructed to try to live this month like it was their last in their current city, serving as a time scarcity manipulation. For this condition, the time "counted down" (e.g., the first week, they read that they had 30 days left, the second week they read that they had 21 days left, etc.). Participants in the control condition were instructed to simply keep track of their daily activities. Once a week for four weeks, participants wrote about their activities for 8 minutes in response to their condition-specific writing prompt. Participants completed follow-up measures of well-

being at two time points: directly after the intervention had concluded and again two weeks later. With the exception of the first time point, participants also completed the Brief Measure of Psychological Needs (BMPN). With the exception of the consent form, the materials were identical to that of the original authors. Using the same procedure as the original authors, we standardized life satisfaction as measured by the SWLS and the positive emotions and (reverse-scored) negative emotions as measured by the mDES to create a composite well-being variable. Following our pre-registered analysis, we conducted a linear mixed effects model, regressing the composite well-being (WB) variable on condition (LTM vs. control), week (baseline vs. post-intervention vs. 2 weeks post-intervention) and their interaction, with a random effect of subject. We did not observe the predicted interaction between condition and week, $F(2, 378.51) = 0.68$, $p = 0.51$. That is, well-being did not change more over time for the LTM condition as opposed to the control. Moreover, WB did not change at a different rate between the two conditions when we examine only the baseline vs. post-intervention, $F(1,184.03) = 0.01$, $p = 0.92$, or baseline vs. 2-weeks post-intervention, $F(1,210.37) = 0.92$, $p = 0.34$. We also obtained the R code used by the original authors for their analyses. They ran a linear mixed effects model using the R "nlme" package. We repeated their analyses using our replication data. In this analysis, condition did not significantly predict well-being: $\gamma11=-0.01$, SE=0.08, $t(442) =-0.13$, $p = 0.90$, $d = -0.03$.The authors of the original study excluded participants who did not have complete well-being data at each of the three time-points (baseline, post-intervention, and 2 weeks post-intervention). As such, we conducted an identical version of the linear mixed effects model but using a replication dataset that also excluded participants who did not complete well-being measures at all

three time-points. This dataset included 132 participants. As with the model conducted using the full dataset, condition did not significantly predict well-being in the post-exclusions data set: $\gamma 11 = -0.17$, SE=0.13, $t(130) = -1.26$, $p = 0.21$, d = -0.45

Lee, Baumgartner, & Winterich, (2018) Study 4

Original: Does a consumer's attitudes towards luxury brands differ when the observed consumption is enabled by unearned (vs. earned) resources, and does a consumer's fairness values play a role in moderating this relationship? Participants were 251 undergraduates assigned to one of six conditions in a 3 (consumption resources: earned, unearned, or unspecified) X 2 (fairness values: low vs. high) between-subjects design. First, participants completed a writing task that manipulated their fairness values. Participants in the high fairness values condition were shown fairness vignettes drawn from moral foundation theory and then wrote one essay about how the examples violated the principle of fairness and a second essay about why fairness is important to them. Participants in the low fairness values condition were shown vignettes about everyday situations and wrote two essays about their everyday activities. Next, participants read a news article about college students' consumption of luxury fashion items. This news article manipulated consumption resources by having people featured in the story mention that students worked part time and saved money to make their purchases (earned), that students receive money from their parents without working (unearned) or making no mention of the source of resources. Finally, participants completed measures for brand attitude, prestige perceptions, attitudes towards the target consumers, dis-association motives, and envy. An ANOVA predicting brand attitude was performed using consumption resources, fairness, and their interaction as independent variables.

There was a significant effect of consumption resources ($F(2, 245) = 3.85$, $p < .05$, $\eta2 = .029$), and a significant interaction ($F(2, 245) = 4.04$, $p < .05$, $\eta2 = .031$). This focal interaction was also significant when the earned and unspecified conditions were combined ($F(1, 247) = 6.47$, $p = .01$, $\eta2 = .025$). When fairness was salient, unearned resources led to lower brand attitude compared to earned or unspecified resources ($M = 5.53$ vs. $M = 6.50$, $t(247) = 2.81$, $p = .005$, Cohen's $d = -0.51$) but not when fairness was not salient ($M = 6.65$ vs. $M = 6.38$, $t(247) = -0.78$, $p > .44$, Cohen's $d = 0.16$).

Replication: We recruited 642 workers from Amazon Mechanical Turk to complete this study. Thirty-eight participants failed an instructional manipulation check ("Indicate Disagree for this item"), leaving 604 participants ($M_{age} = 38.20$, $SD_{age} = 12.52$, 246 males, 353 females, 5 non-binary) in all analyses reported below. We randomly assigned participants to one of six conditions in a 3 (consumption resources: unearned vs. earned vs. unspecified) by 2 (fairness values: low vs. high) between-subjects design. We first instructed participants to complete a writing task as a manipulation of fairness values. Next, to manipulate consumption resources, participants were presented with a news article about a recent study on college students' consumption of luxury accessories from the following brands: Gucci, Burberry, Prada, Louis Vuitton, Hugo Boss, Armani, Coach and Michael Kors. In the earned [unearned] consumption condition, it was stated that students purchase luxury items using their own [parents'] money. In the unspecified condition, no information about the source of consumption resources was mentioned. After reading the news article, participants were asked to indicate their attitudes toward the brands mentioned in the news article (1 = Negative/Bad/Dislike to 9 = Positive/Good/Like, $\alpha = 0.97$) and perceptions of the consumer's prestige (4 items, e.g.,

"The students mentioned in the news article have high status", α = 0.83). Additional

measures that were used to test alternative mechanisms included attitude toward the

target consumers (1 = Negative/Bad/Dislike to 9 = Positive/Good/Like, α = 0.97),

(dis)association motives (3 items, e.g., "I want to avoid these brands due to their

association with the students mentioned in the article", α = 0.63), and envy (2 items, e.g.,

"I am envious of the college students described in the article", $r = 0.87$, $p < .001$). Then

participants were asked a categorical manipulation check question regarding consumption

resources (more than 89% correct responses for all conditions). Finally, they provided

background information, including gender, age, familiarity with the brands, liking toward

the brands, ownership of any brands, subjective SES, and family social class.

We conducted an ANOVA[1] using consumption resources, fairness, and their interaction

as independent variables and brand attitude as the dependent variable. Only the effect of

consumption resources was significant, $F(2, 598) = 4.28$, $p = .014$, ηp2 = .014, η2 = .014;

its interaction with fairness was NS, $F(2, 598) = 1.05$, $p = .351$, ηp2 = .003, η2 = .003;

the effect of fairness was NS, $F(1, 598) = 1.01$, $p = .315$, ηp2 = .002, η2 = .002. The two-

way interaction remained NS when each background measure (e.g., familiarity with the

brands) was included as a control variable. When fairness was salient, attitude toward the

brands in the unearned condition (M = 4.74, SD = 2.03) was lower than the earned

condition (M = 5.52, SD = 1.97), t(281) = 2.61, p = .029, but it was not different from the

unspecified condition (M = 5.35, SD =2.18), t(281)1 = 2.02, p = .134; the earned and

unspecified conditions did not differ, t(281) = 0.58, p = 1.000. When fairness was not

---

[1] The primary analyses, using R, report the post hoc comparisons as Bonferroni-adjusted *t*-tests. The secondary analyses, using SPSS, report the post hoc comparisons as ANOVA contrasts using pooled *df*.

salient, attitude in the unspecified (M = 5.24, SD = 2.35), earned (M = 5.67, SD = 1.90), and unearned conditions (M = 5.23, SD = 2.07) did not differ from each other (unspecified vs. earned, t(317) = 1.47, p = .428; unspecified vs. unearned, t(317) = 0.03, p = 1.000, earned vs. unearned, t(317) = 1.52, p = .387). In addition, salient fairness (vs. neutral) did not lower brand attitude in all consumption resources conditions (unspecified: t(201) = 0.37, p = .713; earned: t(199) = 0.53, p = .595; unearned: t(198) = 1.67, p = .097).When the earned and unspecified conditions were combined, the resource X fairness interaction on brand attitude was still NS, F(1, 600) = 1.76, p = .185, ηp2 = .003, η2 = .003. Simple effect analyses showed that unearned resources led to lower brand attitude compared to earned or unspecified resources (M = 4.74 vs. M = 5.44, t(600) = 2.61, p = .009) when fairness was salient, but not when fairness was not salient (M = 5.23 vs. M = 5.45, t(600) = 0.89, p = .371; Figure 1). The resource X fairness interaction was NS when the unearned and unspecified conditions were combined, F(1, 600) = 0.005, p = .946, ηp2< .001, η2 < .001. *Reduced model.* We identified a reduced model that we believe is a better test of the primary hypothesis in Study 4 of Lee, Baumgartner, & Winterich, 2018. We were able to calculate estimated results of this reduced model for the original study using information provided in Table 3 of Lee, Baumgartner, & Winterich, 2018. We also ran this reduced model using the replication data. Specifically, we conducted an ANOVA using consumption resources, fairness, and their interaction as independent variables and brand attitude as the dependent variable but dropped all participants in the unspecified condition. This results in a reduced model: resources (earned vs. unearned) x fairness (salient vs. not salient). The resource x fairness interaction in this reduced model was not significant, F(1, 397) = 0.73, p = .393, ηp2 =

0.002. *Mediation analysis.* Following the analysis strategy in the original paper and the preregistration, we proceeded to test for moderated mediation via consumer prestige perceptions using bootstrapping (Model 8 in PROCESS; Hayes, 2012). Consumption resource was coded as: unspecified = 1, earned = 1, unearned = -2. The interaction of consumption resource and fairness on prestige perception was NS, $\beta = 0.04$, $t(600) = 0.50$, $p = .615$. Prestige perceptions positively affected brand attitude, $\beta = 0.62$, $t(599) = 10.24$, $p < .001$. The indirect effect of consumption resources on brand attitude via prestige perceptions was NS, both when fairness was salient (0.07, 95% CI = [-0.0012, 0.14]) and not salient (0.04, 95% CI = [-0.02, 0.11]). The index of moderated mediation was also NS (0.02, 95% CI = [-0.07, 0.12]), suggesting that the moderated mediation model was not supported. We also conducted additional moderated mediation analyses with general attitude toward the consumer, (dis)association motives, and envy as mediators, respectively. For all three models, the index of the moderated mediation was NS.

Mani, Mullainathan, Shafir, & Zhao, (2013) Study 4

Original: Does priming poverty reduce fluid intelligence and executive control? Does this depend on participants' actual income? Participants were 96 adults recruited from a New Jersey shopping mall who were assigned to one of two experimental conditions in this two-cell design. Participants' income was measured by dividing self-reported household income by the square root of the size of the household, and then taking a median split of this value to yield a categorical variable. All participants read four hypothetical scenarios relating to a financial problem (e.g., having to pay for car repairs). Participants were told they could take a loan, take a chance and not address the problem,

or pay in full. Participants in the "hard" condition were told the costs of addressing each problem were relatively high (e.g., $1,500), while those in the "easy" condition were told the costs were relatively low (e.g., $150). The amounts in the easy condition were expected to have little effect on cognitive function for neither the high-income nor the low-income participants, but the amounts in the hard condition were expected to evoke financial concerns in only the low-income participants. Participants indicated how they would respond to each of these scenarios before completing the Raven's Progressive Matrices task and a spatial compatibility task. Participants' responses to the Raven's matrices task and spatial compatibility tasks were submitted to separate ANOVAs with condition, income, and their interaction as predictors. These ANOVAs revealed the predicted interactions, such that in the easy conditions, the rich and poor performed similarly, but in the hard conditions, the rich performed better than the poor; for the Raven's matrices, $F(1,92) = 4.04$, $p = 0.04$; for cognitive control: $F(1,92) = 6.66$, $p = 0.01$.

 Replication: We recruited 236 workers from Amazon's Mechanical Turk (this was 4 fewer than our initial target, due to a programming error). After manually reviewing and removing 7 careless responses (e.g., responses that were completely nonsensical or did not attempt to answer the prompt), 229 valid responses remained. Following Study 4 from Mani, Mullainathan, Shafir, & Zhao (2013), a household income-diverse sample was recruited (Median = $55,000, minimum = $5000, maximum = over $200,000). Effective income was computed by dividing household income by the square root of household size. "Rich" and "poor" participants were defined through a median split on this variable. Participants were randomly assigned to one of two conditions in a 2 cell

(financial hardship: hard or easy) between-subjects design. We asked participants to imagine and write about how they would deal with four hypothetical financial scenarios, which varied in whether they described a large (hard) or small (easy) expense. After writing about how they would solve each scenario, participants performed a set of three Raven's Progressive Matrices. As participants responded to four scenarios with three Raven's matrices per scenario, participants answered twelve matrices in total. The dependent variable was the proportion of correct responses to these twelve matrices. Household income and size were measured after responding to all scenarios and matrices. Data were analyzed using a two-way ANOVA, following Mani et al. (2013), wherein Raven's matrices scores were predicted from status as "rich" or "poor", assignment to "easy" or "hard" scenarios, and their interaction. Contrary to prior results (Mani et al., 2013), there was no robust interaction between income and condition ($F(1,224) = 0.17$, $p = .678$). The rich and poor performed similarly in the easy condition ($t(109) = -0.10$, $p = .918$) and in the hard condition ($t(115) = -0.70$, $p = .487$). Condition was insignificant for the rich ($t(114) = -0.16$, $p = .873$) and for the poor ($t(110) = 0.43$, $p = .670$). Again, contrary to prior results, the poor did not perform worse than the rich overall ($F(1,224) = 0.13$, $p = .718$). There was no main effect of condition ($F(1,224) = 0.18$, $p = .670$). Overall, participants in all four cells of the design answered around 30% of matrices correctly, which fell between the accuracy rates reported in Mani et al., (2013; from 20–40% correct across conditions). Including careless responders did not qualitatively change the key study result: there was no robust interaction between income and condition ($F(1,231) = 0.07$, $p = .798$).

Mehta & Zhu, (2016) Study 2

Original: Does being primed with scarcity cause greater product use creativity by leading participants to think beyond a product's typical functionality to solve a problem? Participants were 153 adults recruited online who were assigned to one of three conditions in this three-cell design. Participants were assigned to either a "scarcity" condition, an "abundance" condition, or a control condition. Participants in the scarcity [abundance] condition completed a writing task in which they wrote an essay about growing up with scarce [abundant] resources. Participants in the control condition did not complete any writing task. All participants then completed the candle creativity task, which requires convergent creative thinking to generate a novel use for a product to solve a problem. Specifically, participants were shown a picture of a candle, a pack of matches, and a box of tacks on a table next to a wall and were told to attach the candle to the wall using the objects in the picture, so the burning candle does not drip wax on the table or the floor. Participants were asked to write down the solution to the task. The correct response is to empty the box of tacks, use the tacks to affix the box to the wall, and place the candle in the box. Participants' responses to this task were coded as correct or incorrect. A chi-square showed an effect of resource availability on correct responses, $chi2(2, N = 133) = 10.69$, $p = .005$). A binary logistic regression analysis showed that a higher percentage of participants in the scarcity condition ($M = 44.2\%$) correctly solved the candle problem as compared to those in the abundance ($M = 15.6\%$, $B = 1.46$, standard error [SE] $= .51$, Wald $= 8.07$, $p = .005$) and control ($M = 20.0\%$, $B = 1.15$, SE $= .48$, Wald $= 5.70$, $p = .017$) conditions. No difference was observed between the abundance and control conditions ($B = .31$, SE $= .56$, Wald $= .30$, not significant).

Replication: We recruited 387 workers from MTurk to complete this study (185 women) completed an online. We randomly assigned participants to one of three resource availability conditions (scarcity, abundance, and control) in this 3-cell design. The participants assigned to the scarcity and abundance conditions began the study by completing a writing task manipulation of resource availability. The participants in the control condition proceeded directly to the second task. Next, all participants were presented with the candle task developed by Duncker (1945). The participants were shown a picture containing the following products on a table: a candle, a pack of matches, and a box of tacks, all of which were next to a wall. Participants were directed to figure out how to attach the candle to the wall by using only the objects on the table, so that the candle burns properly and does not drip wax on the table. After writing down the solution to this task, the participants were asked whether they had knowledge of the candle task and its solution beforehand. Participants who indicated that they knew of the task and its solution beforehand were excluded from analysis. Finally, participants answered demographic questions.

After manually reviewing and removing 8 careless responses (e.g., responses that were completely nonsensical or did not attempt to answer the prompt), 381 valid responses remained. We excluded data from 61 participants who indicated having knowledge about the candle task and its solution beforehand. The remaining 319 responses were coded as correct or incorrect in line with previous literature; for a solution to be considered correct, responses had to include the use of the box of tacks as a candle holder (Maddux and Galinsky, 2009). Ninety-nine of the 327 participants (31%) correctly solved the problem. A chi-square test revealed no significant effect of resource availability on the correctness

of the solutions (X^2 (2, N = 319) = 2.224, p = .329). We next conducted a binary logistic regression analysis to assess the differences between the conditions. The results showed that there were no significant differences between the scarcity condition (M=27%) and the abundance condition (M = 29%, B = -.02, [SE] = .06, Wald = .063, p = .802), and that there was no significant difference between the scarcity condition and the control condition (M = 36%, B = -.09, [SE] = .06, Wald = 1.90, p = .174.) Furthermore, we observed no difference between the abundance and control conditions (B = -.07, [SE] = .06, Wald = 1.2, p = .278). Including careless responders did not qualitatively change the key study result: there was still no significant difference between the scarcity and control condition (M = 36%, B = -.10, [SE] = .06, Wald = 2.422, p = .120).

Monga, May, & Bagchi, (2017) Study 4

Original: Are implied wage rates higher in a time elicitation procedure than in a money elicitation procedure for the near future, but not for the distant future? Participants were 189 adults recruited from MTurk who were assigned to one of four conditions in a 2 (elicitation procedure: time or money) X 2 (timing: near future or distant future) fully between-subjects design. All participants were asked to consider participating in a future MTurk survey in return for compensation. Participants in the near future condition were told that the upcoming study would start at 8:00am on a Saturday in 3 days (the study took place on a Wednesday) and participants in the distant future condition were told that the upcoming survey would start at 8:00am on a Saturday in 24 days. Participants in the money elicitation condition (MEL) were told the survey would take four hours and asked to indicate the minimum amount of money they would require to complete the survey, while those in the time elicitation condition (TEL) were told they would be paid $24 and

asked to indicate the maximum amount of time they would be willing to work on the survey. These values were converted to a wage rate per hour by dividing $24 by the number of hours the participant would be willing to work. Participants hourly wage rates were submitted to an ANOVA with elicitation procedure, timing, and their interaction as predictors. The ANOVA returned a main effect of elicitation procedure, a main effect of timing, and a significant two-way interaction between procedure and timing ($F(1, 175) = 8.60$, $p < .01$), such that the wage-rate asymmetry was stronger in the near future ($M_{TEL-near\ future} = \$11.77$, $SD = \$7.75$; $M_{MEL-near\ future} = \$4.30$, $SD = \$2.54$; $F(1, 175) = 50.83$, $p < .01$, Cohen's $d = 1.29$) than in the distant future, ($M_{TEL-distant\ future} = \$7.72$, $SD = \$3.55$; $M_{MEL-distant\ future} = \$4.52$, $SD = \$3.10$; $F(1, 175) = 10.03$, $p < .01$, Cohen's $d = .96$).

Replication: We recruited 474 workers from MTurk and randomly assigned them to one of two conditions in this 2 (elicitation procedure: time or money) X 2 (timing: near future or distant future) between-subjects design. Following the original paper's guidelines, we dropped data from 7 participants who provided responses three or more standard deviations beyond the mean. We analyzed the remaining 467 participants (57% female, Mage=35.7). All participants were asked to consider participating in a future MTurk survey in return for compensation. Participants in the near future condition were told that the upcoming study would start at 8:00am on a Saturday in 3 days (the study took place on a Wednesday) and participants in the distant future condition were told that the upcoming survey would start at 8:00am on a Saturday in 24 days. Participants in the money elicitation condition (MEL) were told the survey would take four hours and asked to indicate the minimum amount of money they would require to complete the survey, while those in the time elicitation condition (TEL) were told they would be paid $24 and

asked to indicate the maximum amount of time they would be willing to work on the survey. These values were converted to a wage rate per hour by dividing $24 by the number of hours the participant would be willing to work. Next we conducted an ANOVA predicting the wage rate as the dependent variable from an indicator variable for wage-elicitation procedure (MEL vs. TEL), an indicator variable for the timing of the activity (near vs. distant), as well as an interaction term. A main effect of procedure emerged, such that the desired wage rate was higher in the TEL (vs.MEL) condition (M_{TEL} = $15.06, SD = $9.54; M_{MEL} = $8.71, SD = $8.21; $F(1, 463)$ = 27.26, $p < .001$, Cohen's d = 0.179). There was no significant difference of the wage rate in the distant versus near future ($M_{near\ future}$ = $12.44, SD = $10.63; $M_{distant\ future}$ = $11.41, SD = $8.14; $F(1, 463)$ =0.32, $p = 0.6$, Cohen's d = .028). Additionally, there was no significant two-way interaction between procedure and timing ($F(1, 463)$ = 0.139, $p = 0.71$; M_{TEL}-Near Future = $15.71, SD = $10.68; M_{MEL}-Near Future = $9.05, SD = $9.49; M_{TEL}-Distant Future = 14.43, SD =8.27; M_{MEL} distant future = 8.38, SD = 6.80).

Plantinga, Krijnen, Zeelenberg, Breugelmans, (2019) Study 3

Original: Do poor people spontaneously consider opportunity cost more often than people with higher incomes, and does this lead to lower opportunity cost neglect? Does reminding people of opportunity costs reduce this gap? Participants were 637 adults recruited from MTurk and assigned to one of two conditions in this two-cell design. All participants were asked to consider an attractive product (a movie ticket for $8.50) and asked whether they would buy the product. Participants in the control condition were not reminded of the opportunity cost (the non-buying options was phrased as "not buying the product") while those in the experimental condition were reminded of the opportunity

cost (the non-buying option was phrased as "keeping the money for other purchases.")

Participants then were asked to list alternative things they could do with the money and to

indicate how difficult it was to generate the alternatives. Finally, participants reported

their income and other demographics. A chi-squared test revealed that participants in the

experimental condition were less likely to buy the concert ticket than those in the control

condition X^2, (2, n = 642) = 18.78, p < .001. However, a logistic regression to test for the

interaction between condition and effective income (measured in increments of $10,000

and centered for each increment) did not reveal a significant effect (b = -.12, se = .08),

indicating that reminders of opportunity cost did not affect the wealthy more than those

with lower incomes. Income did not predict ease of generating alternative uses for

money, but it was positively correlated with number of alternative uses, r(594) = .05,p=

.244, 95% CI [.03, .13].

Replication: We recruited 1613 workers from Amazon's Mechanical Turk to complete

this study. Manual review identified 59 careless responses (e.g, nonsensical responses or

responses that did not attempt to address the study prompt), resulting in total valid sample

size of 1564. Participants were randomly assigned to one of two conditions in this 2-cell

(opportunity cost or control) design. In the Opportunity Cost condition participants chose

between buying a movie ticket, described in a brief preceding vignette, or keeping the

money for other purchases. In the Control condition, participants simply chose between

buying the movie ticket versus not. Following their decision to buy, all participants were

asked to generate alternative uses of the money, and to rate how difficult it was to do so.

At the end of the survey, participants provided demographic information, including

objective and subjective measures of wealth. Participants were distributed across income

quintiles in the U.S., with 23.8% in the lowest quintile, and 21.7, 28.4, 18.9, and 7.2 in the 2nd, 3rd, 4th, and 5th, respectively. Approximately one-tenth (10.4%) of participants were below the U.S. poverty line. We conducted a logistic regression on decision to buy (coded as '1' vs. '0'), with condition (Control vs. Opportunity Cost), centered effective income (in $10,000), and their interaction, to test the primary hypothesis that effect of reminding participants of opportunity costs should be smaller for lower income participants. The interaction was not reliable (OR: 1.02, z = .49, 95% CI [.94, 1.11]), nor was the effect of income (OR: 1.01, z = .25, 95% CI [.95, 1.07]). However, even controlling for income and the interaction effect, the effect of condition was reliable (OR: .60, z = -4.51, 95% CI [.48, .75]), with fewer decisions to buy in the Opportunity Cost condition compared to Control (65.84% vs. 76.61%, $X2(1) = 22.31$, p <. 001). There was no significant correlation between effective income and willingness to buy (r(1563) = .03, z = 1.20, p = .23, CI [-.02, .08]), nor difficulty (r(1563) = .03, z = 1.02, p = .31, CI [-.02, .08]) or number (r(1563) = .03, z = 1.01, p = .31, CI [-.02, .08]) of generated alternatives. Including careless responses (N = 49) did not qualitatively change the key study result: the interaction was not reliable (OR: 1.03, z = .80, 95% CI [.95, 1.12]), nor was the effect of income (OR: 1.00, z = .25, 95% CI [.95, 1.07]). However, even controlling for income and the interaction effect, the effect of condition was reliable (OR: .59, z = -4.58, 95% CI [.47, .74]), with fewer decisions to buy in the Opportunity Cost condition compared to Control (65.25% vs. 76,33%, $X2(1) = 22.72$, p <. 001).

Roux, Goldsmith, & Bonezzi (2015) Study 4

Original: Do reminders of scarcity cause selfish behavior to a greater extent in people with low social value orientation (SVO) relative to those with high SVO? Participants

were 157 adults recruited from MTurk who were assigned to one of two conditions in this two-cell design. Participants in the scarcity condition were sequentially shown five resources (gasoline, sugar, water, wheat, and electricity) and asked to list three things three wouldn't be able to do if those resources were unavailable. Participants in the control condition were sequentially shown the same five resources and asked to list three things they could do with each resource. Participants then completed a dictator game in which they imagined allocating $5 in $1 increments between themselves and another anonymous individual. Next, participants completed an unrelated study for five minutes before completing a six-item measure of Social Value Orientation. A composite measure for SVO was computed such that lower scores indicate a pro-self orientation and higher scores indicate a pro-social orientation. The amount participants allocated to the imaginary other player in the dictator game was regressed on a dummy code for scarcity condition, a continuous variable for SVO, and their interaction. The results showed a main effect of scarcity condition, beta = -.38, SE = .14, t = 2.83, p = .005, a significant effect of SVO, beta = .02, SE = .005, t = 4.07, p < .001, and a significant interaction, beta = .01, SE = .005, t = 2.12, p = .04.

Replication: We recruited 422 workers from Amazon Mechanical Turk to participate in this study. After manually reviewing and removing 8 careless responses (e.g., responses that were completely nonsensical or did not attempt to answer the prompt), 414 valid responses remained. All participants viewed 5 different resources (gasoline, sugar, water, wheat and electricity) sequentially. We randomly assigned participants to one of two conditions in this two-cell (scarcity or control) design. To induce thoughts of scarcity, some participants were randomly assigned to list three things they could not do if each

resource was unavailable. In the control condition, participants listed three things they could do with each resource. Immediately after completing the experimental task, participants played a simulated dictator game. We asked participants to imagine they were playing a real but anonymous individual, and they had to allocate $5 between themselves and this other player. They responded using a slider anchored at sending $0 to $5 with $1 increments. After completing this task, participants completed an approximately 5-minute filler task (BFI-2 inventory) and then completed the measure of the moderator, the social value orientation scale. An ANOVA testing for the effect of the scarcity manipulation on allocation behavior revealed no significant effects of the scarcity manipulation (M = $1.65, SD = 1.18) in comparison to the control condition (M = $1.48, SD = 1.17; F (1, 412) = 0.56, p = .45; η_p^2 = .001). We scored SVO according to the procedures outlined by Murphy et al. (2011) for each participant. Before analyzing the interaction effect, we confirmed there is no significant effect of condition on SVO because it was measured after the manipulation (p = .08). To test for the primary hypothesis, an attenuated moderation of scarcity by SVO, we regressed the amount of money allocated to the other player in the dictator game on scarcity (dummy coded: -1 = control, 1 = scarcity), SVO and their interaction. The results from a type 3 ANOVA indicate a significant main effect of SVO (β = .43, SE = .04, t(410) = 9.73, $p < .001$; η_p^2 = .19), but no main effect of the manipulation manipulation (β = 0.001, SE = .04, t(410) = -0.13, p = .90; η_p^2 = .00). There was also no significant interaction (β = .01, SE = .04, t(410) = 0.16, p = .88; η_p^2 = .00. Because there was no significant interaction effect, we did not perform the floodlight analysis or check the simple slopes of SVO by condition. Including careless responders did not qualitatively change the key study result: there was

still no significant interaction between SVO and the manipulation, ($\beta = .02$, SE = .04, $t(418) = 0.49$, $p = .62$; $\eta p2 = .00$).

Shah, Zhao, Mullainathan, Shafir (2018) Study 3

Original: Do thoughts triggered by financial concerns intrude into consciousness for poorer individuals more often than for wealthier individuals? Participants were 573 adults (source not stated) who were assigned to one of two conditions in this two-cell design. Income was measured using a scale with $10k ranges from below $10k to $150k and above coded as the midpoint of the scale (except $150k). This value was divided by the square root of the household size. This new value was then split at the median and used as a measured categorical independent variable. All participants were told to let their minds wander freely for three minutes and to talk out loud continuously, saying whatever crossed their mind. Next, participants were asked to let their mind wander again for an additional three minutes, but to not think about an aspect of their driving. Participants in the control condition were told not to think about how much they drive, while participants in the cost condition were told not to think about how much driving costs them. Anytime participants had one of these forbidden thoughts cross their mind, they were instructed to press a red button on the screen that says, "I thought of it." An ANOVA predicting number of intrusions from categorical variables indicating condition, income, and their interaction returned a significant interaction $F(1, 564) = 6.99$, $p < .01$, partial eta2 = .01. The interaction was also significant when income was treated continuously, $b = -3.43$, $t(564) = 2.38$, $p < .02$. When asked to not think about how much they drove, the number of intrusions for lower-income participants (5.77 [4.86, 6.68]) did not significantly differ from intrusions for higher-income participants (6.87 [5.74, 8.01]).

Yet when asked to not think about how much they spend on driving, lower-income participants reported significantly more intrusions (6.63 [5.48, 7.77]) than did higher-income participants (5.01 [4.17, 5.85]).

Replication: We recruited 1436 workers participants from MTurk to complete this study (664 females, 764 males; median household size: 2 people; median household income: $55,000 [*SD* = $38,190]). We randomly assigned participants to one of two conditions in this two-cell (cost or time) design. We first gave participants the instructions to let their minds wander freely for three minutes. Following the first mind-wandering phase, we asked participants to let their minds wander for an additional 3 minutes. This time, we instructed them to not think about some aspect of their driving. Specifically, some participants were told to not think about how much they drive, while other participants were told to not think about how much driving costs them. We instructed participants to click the button each time they had an intrusive thought. After removing 102 observations with missing data, there were 1334 participants available for analyses. When asked to not think about how much they drove, the number of intrusions for lower-income participants (M = 7.84, CI = [6.88, 8.80]) did not significantly differ from intrusions for higher-income participants (*M* = 8.47, CI = [7.27, 9.68]), *t*(623) = -0.81, p = .42. When asked to not think about how much they spend on driving, lower-income participants reported no more intrusions (*M* = 7.62, CI = [6.56, 8.69]) than did higher-income participants (M = 7.19, CI = [6.22, 8.15]), *t*(654) = 0.59, *p* = .55; interaction between income and condition *F*(1, 1330) = 1.00, *p* = .32 partial eta-squared(np2) =.001. The interaction was also not significant when income was treated continuously, *b* = -0.05, *t*(1330) = -0.078, *p* = .94.

Shah, Shafir, & Mullainathan (2015) Study 6

Original: Does scarcity override context effects, causing poorer people to view losses in more absolute, as opposed to relative, terms than wealthier people? Participants were 74 adults recruited from MTurk who were assigned to one of four conditions in this 2 (scarcity condition: poor or rich) X 2 (account condition: large or small) fully between-subjects design. All participants completed five rounds of a Family Feud game in which they tried to guess the top 5 answers to a number of different prompts. Participants received 1 point per correct answer that they guessed. Participants in the *poor* scarcity condition were given 15 seconds per round to guess answers to the prompt, while those in the *rich* scarcity condition were given 50 seconds per round to guess. Next, participants in the *small* account condition were asked to consider their time budget for one round of trivia, while those in the *large* account condition were asked to consider their time budget for the overall game. Finally, participants rated how expensive or costly they believed a 10 second loss of time would be on an 11-point scale ranging from 1 to 11. A 2-way ANOVA predicting Family Feud scores from factor variables indicating scarcity condition and account condition returned a significant interaction effect, $F(1, 69) = 5.16$, $p < .05$, $\eta p2 = .07$. Time-rich participants rated the loss as more expensive when they thought about a small account ($M = 8.31$, 95% CI = [7.78, 8.84]) than when they thought about a large account ($M = 6.50$, 95% CI = [5.42, 7.58]), whereas time-poor participants' evaluations did not differ between the small-account condition ($M = 8.33$, 95% CI = [7.14, 9.52]) and the large account condition ($M = 8.83$, 95% CI = [7.97, 9.69]).

Replication: We recruited 209 Amazon Mechanical Turk workers to complete this study. Their mean age was 37.6 years ($SD = 19.8$), 46.9% identified as female. We randomly assigned participants to one of four conditions in this 2 (scarcity condition: poor or rich)

X 2 (account condition: large or small) fully between-subjects design. Participants played the trivia game "Family Feud" in which they were to guess the most popular responses from a sample of 100 previous respondents on association tasks (e.g. name things you would take on a picnic). The game consisted of five rounds, for each round, participants received one point for each correct guess of the top five most popular answers. We randomly assigned participants to one of the two scarcity conditions, i.e. were primed to think about their time budget for one round vs. the entire game. After playing Family Feud, participants were randomly assigned to have a small or large account. Participants then indicated how expensive it would have felt if they had lost 10 seconds from their time budget due to a computer glitch on a scale from 1 (not expensive at all) to 11 (very expensive). There was no significant interaction between time-scarcity and account manipulation. Participants in the time-scarce condition did not rate the 10s loss as more costly in the small account ($M = 8.69$, 95% CI = [8.03; 9.35]) than in the large account condition ($M = 8.98$, 95% CI = [8.41; 9.55]). Likewise, participants in the time-rich condition did not exhibit a framing effect, as participants did not rate the 10s loss as more costly in the small account ($M = 7.34$, 95% CI = [6.55; 8.13]) than in the large account condition ($M = 6.73$, 95% CI = [5.97; 7.49]). A total of 57% indicated the correct time budget when asked how much time they were given for the first round (in the small account condition) or for the game altogether (in the large account condition). The proportion of correct responses were evenly distributed across all four conditions. When analyzing the subset of participants who recalled the time budgets correctly, the results remained unchanged. A 2 (scarcity condition) $\times$ 2 (account condition) analysis of variance revealed no significant interaction, $F(1, 205) = 1.68$, $p = .20$.

Tully, Hershfield & Meyvis (2015) Study 5

Original: Does scarcity cause consumers to prefer material goods to experiential goods? Participants were 407 workers recruited from MTurk who were assigned to one of two conditions in this two-cell design. All participants were first asked to complete a writing task. Participants in the *scarcity* condition were asked to write about financial constraints in their lives, while those in the *control* condition were asked to list 10 facts they knew to be true. Next, participants read about five scenarios in which they had decided to spend money for a specific purpose (e.g., getting in shape) and could buy either a material good (e.g., buying gym equipment) or an experiential good (e.g., buying a gym membership). For each scenario participants indicated their preference between the two-options on a 7-point scale. Responses to this item were re-coded such that higher scores indicated a greater preference for the material option. Next, participants were asked to indicate the reasons for each choice from a list of possible reasons and a free-response item. Participants then completed a manipulation check, four mood items from the PANAS, demographic questions, and an attention check. A repeated-measures ANOVA with item pair as a within-subjects factor and financial constraints as a between-subjects factor revealed a significant effect of financial constraints. As predicted, participants who were asked to consider their financial constraints were more likely to prefer the material options than were participants in the control condition ($F(1, 375) = 10.02$, p = .002). This effect again did not differ across pairs ($F(4, 1500)= 1.24$, *NS*).

Replication: We collected data from 1,012 workers recruited on MTurk, which yielded a final sample size of 712 participants after our pre-registered exclusion criteria were applied. After manually reviewing and removing 24 careless responses (e.g., responses

that were completely nonsensical or did not attempt to answer the prompt), 688 valid

responses remained. We randomly assigned participants to one of two conditions in this

two-cell (scarcity or control) design. All participants were first asked to complete a

writing task. Participants in the *scarcity* condition were asked to write about financial

constraints in their lives, while those in the *control* condition were asked to list 10 facts

they knew to be true. Next, participants read about five scenarios in which they had

decided to spend money for a specific purpose (e.g., getting in shape) and could buy

either a material good (e.g., buying gym equipment) or an experiential good (e.g., buying

a gym membership). For each scenario participants indicated their preference between the

two-options on a 7-point scale. Responses to this item were re-coded such that higher

scores indicated a greater preference for the material option. Next, participants were

asked to indicate the reasons for each choice from a list of possible reasons and a free-

response item. Participants then completed a manipulation check, four mood items from

the PANAS, demographic questions, and an attention check. First, as a test of the

manipulation, we compared participants' self-reported responses of the extent to which

they thought about financial constraints and the extent to which they felt financially

constrained. We found that participants in the financial constraint condition both reported

thinking more about financial constraints, $M_{scarcity} = 5.53$, $SE = 1.81$, $M_{control} = 4.04$, $SE =$

2.14, $t = -9.89$, $p < .001$, and feeling more financially constrained, $M_{scarcity} = 5.37$, $SE =$

1.69, $M_{control} = 4.35$, $SE = 2.00$, $t = -7.29$, $p < .001$. Next we submitted participants ratings

for each item pair to a repeated-measures ANOVA with pair as a within-subjects factor

and financial constraints as a between-subjects factor revealed a significant effect of

financial constraints. Consistent with Tully, Hershfield, and Meyvis (2015) participants

who were asked to consider their financial constraints were more likely to prefer the material options than were participants in the control condition $(F(1, 686) = 12.84, p = .0004$, partial-eta$^2 = .018)$. As in the original study, this effect differed across pairs $(F(4, 686) = 5.59, p < .001$, partial-eta$^2 = .008)$.

Zhu & Ratner, (2015) Study 2

Original: Does perceived scarcity affect preferences by polarizing relative liking for favorite versus non-favorite alternatives and increase choice share for the favorite option? Participants were 200 adults recruited online who were assigned to one of two conditions in this two-cell design. Participants were assigned to an *abundant* or *scarce* supply-level condition. All participants were told the researchers were studying vegetable preferences and saw eight pieces of four types of vegetables: baby carrots, cherry tomatoes, broccoli florets, and cauliflower florets. In the scarce condition, the vegetables were presented in large 32 oz. food containers and in the abundance condition, the vegetables were presented in small 8 oz. food containers. Participants ranked the four vegetables according to preference and rated how much they liked each vegetable on a 101-point sliding scale. Next, participants were told they could take six pieces of vegetables for a snack and asked to indicate how many of each type they would like. Finally, participants rated how popular they thought each vegetable would be with other participants. To test the primary hypothesis, the researchers calculated the difference between the liking rating of the favorite vegetable and the average of the liking ratings of the non-favorite vegetables. An ANOVA predicting the liking difference from perceived supply level returned a significant main effect $(F(1, 198) = 3.80, p = .05)$. The relative liking of the favorite increased when the supply level of each item in the choice set was perceived as

scarce (M = 37.01, SD = 21.09) compared with abundant (M = 31.10, SD = 21.71). To test for whether scarcity polarized choice share for the favorite option, the researchers divided the number of pieces chosen for the favorite vegetable by the total pieces of vegetables each participant was allowed to choose. A nonparametric Mann–Whitney test for two independent samples to compare choice share of the favorite in the scarce and abundant conditions revealed that participants incorporated a higher proportion of their favorite vegetable when they perceived the supply of each item as scarce (M = 56.29%, SD = 24.35%) than abundant (M = 49.02%, SD = 21.60%), $Z_{Mann–Whitney}$ = –2.20, p = .028.

Replication: We recruited 595 workers from Amazon's Mechanical Turk and randomly assigned them to one of two conditions in this two-cell (abundant or scarce) design. We showed participants the corresponding photograph for each condition with the text:

> "We are interested in studying people's food preference. Please find four types of vegetables (pre-washed, ready-to-eat) as pictured. Imagine that you decided to take six pieces of the vegetables provided to create your own snack mix. You can repeat any of the types as often as you like."

Participants were then asked to (a) rank order the four vegetable types (carrots, baby tomatoes, broccoli, cauliflower) in order of their preference (b) indicate on a sliding scale from 0-100 how much they like each vegetable, with the instruction "Please indicate how much you like each vegetable on a scale from 0 = not at all to 100 = very much" (c) choose six pieces of vegetable to take, with the instructions: "Which six pieces would you pick? Please indicate the number of pieces you would pick from each type. Simply write down "0" for the categories that you decide not to take" (d) indicate on a sliding

scale from 0-100 how popular they perceive each vegetable to be among study

participants, with the instruction "Please indicate how popular you think each vegetable

is with participants of this study (0 = not at all popular, 100 = very popular)." Finally,

participants reported their gender, age, educational attainment, race and ethnicity, total

family income for 2017, and 5-digit zip code. Following Zhu and Ratner (2015) we

computed relative liking of favorite by computing the difference between participants' 0-

100 rating of their ranked-choice favorite vegetable and their mean 0-100 rating of their

three non-favorites. An ANOVA (perceived supply level: abundant vs scarce) on the

liking difference ratings revealed no significant difference between the scarce (M =

27.90, SD = 19.80) and abundant (M = 27.83, SD = 20.91) conditions, $F(1,593) = 0.002$ p

$= 0.97$ d $= 0.003$. Following Zhu and Ratner (2015), to analyze the proportion of the

most preferred vegetable included in participants' choices, we computed the number of

pieces of their favorite divided by the total chosen, which was always six by design. We

performed a Mann-Whitney test for two independent samples to compare choice share of

the favorite in the scarce and abundant conditions. Results suggested that there was no

significant difference between the scarce (M = 0.48, SD = 0.22) and abundant (M = 0.50,

SD = 0.24) conditions on the proportion of favorites chosen, $Z_{Mann–Whitney} = 1.04$, p =

0.30. Finally, following Zhu and Ratner (2015), we computed for each participant the

number of different kinds of vegetables chosen (i.e., minimum 1, maximum 4), and

performed a Mann-Whitney test for two independent samples to compare the number

chosen between the scarce and abundant conditions. Results suggested that there was no

significant difference between the scarce (M = 2.92, SD = 0.96) and abundant (M = 2.86,

SD = 1.04) conditions on the number of different vegetables chosen, $Z_{\text{Mann–Whitney}}$ = -0.22,

p = 0.83.

SI References

1. Abeyta, A. A., Routledge, C., Kersten, M., & Cox, C. R. (2017). The existential cost of economic insecurity: Threatened financial security undercuts meaning. *The Journal of social psychology*, *157*(6), 692-702.

2. Durante, K. M., Griskevicius, V., Redden, J. P., & Edward White, A. (2015). Spending on daughters versus sons in economic recessions. *Journal of Consumer Research*, *42*(3), 435-457.

3. Emery, L. F., Walsh, C., & Slotter, E. B. (2015). Knowing who you are and adding to it: Reduced self-concept clarity predicts reduced self-expansion. *Social Psychological and Personality Science*, *6*(3), 259-266.

4. Fernbach, P. M., Kan, C., & Lynch Jr, J. G. (2015). Squeezed: Coping with constraint through efficiency and prioritization. *Journal of Consumer Research*, *41*(5), 1204-1227.

5. Kristofferson, K., McFerran, B., Morales, A. C., & Dahl, D. W. (2017). The dark side of scarcity promotions: how exposure to limited-quantity promotions can induce aggression. *Journal of Consumer Research*, *43*(5), 683-706.

6. Layous, K., Kurtz, J., Chancellor, J., & Lyubomirsky, S. (2018). Reframing the ordinary: Imagining time as scarce increases well-being. *The Journal of Positive Psychology*, *13*(3), 301-308.

7. Lee, S., Baumgartner, H., & Winterich, K. P. (2018). Did they earn it? Observing unearned luxury consumption decreases brand attitude when observers value fairness. *Journal of Consumer Psychology*, *28*(3), 412-436.

8. Mehta, R., & Zhu, M. (2016). Creating when you have less: The impact of resource scarcity on product use creativity. *Journal of Consumer Research*, *42*(5), 767-782.

9. Monga, A., May, F., & Bagchi, R. (2017). Eliciting time versus money: time scarcity underlies asymmetric wage rates. *Journal of consumer research*, *44*(4), 833-852.

10. Plantinga, A., Krijnen, J. M., Zeelenberg, M., & Breugelmans, S. M. (2018). Evidence for opportunity cost neglect in the poor. *Journal of behavioral decision making*, *31*(1), 65-73.

11. Roux, C., Goldsmith, K., & Bonezzi, A. (2015). On the psychology of scarcity: When reminders of resource scarcity promote selfish (and generous) behavior. *Journal of consumer research*, *42*(4), 615-631.

12. Shah, A. K., Zhao, J., Mullainathan, S., & Shafir, E. (2018). Money in the mental lives of the poor. *Social Cognition*, *36*(1), 4-19.

13. Zhu, M., & Ratner, R. K. (2015). Scarcity polarizes preferences: The impact on choice among multiple items in a product class. *Journal of Marketing Research*, *52*(1), 13-26.

14. Champely, S. (2020). *pwr: Basic Functions for Power Analysis.* (R package version 1.3-0) [Computer software]. https://CRAN.R-project.org/package=pwr

15. Cohen, B. H. (2002). Calculating a Factorial ANOVA From Means and Standard Deviations. *Understanding Statistics*, *1*(3), 191-203 https://doi.org/10.1207/S15328031US0103_04

16. Hughes, J. (2017). *Simulating Power with the paramtest Package*. https://cran.r-project.org/web/packages/paramtest/vignettes/Simulating-Power.html

17. Masur, P. (2020). *Estimating power for structural equation models: Simulations in R*. https://philippmasur.de/2020/03/11/how-to-estimate-power-for-structural-equation-models-simulations-in-r/