

Brief Commentary: Is “4 for \$16” Better than “4 for \$15.30”? Testing the Replicability of the Price Divisibility Effect

AMANDA E. GEISER 
LEIF D. NELSON

Although higher prices generally lead to lower demand, research on the price divisibility effect suggests that consumers are sometimes willing to pay more for a multipack if its price is divisible by the number of units it contains. For instance, a seven-pack of shoelaces might be more appealing at the divisible price of \$14 than at the lower, but non-divisible, price of \$13. The original *Journal of Consumer Research (JCR)* article documenting this phenomenon presented a total of 15 studies, one of which was preregistered and all of which yielded strong evidence. In this article, we report four preregistered replications of three key studies from the original article. Despite our replications' considerably larger sample sizes, none replicated the original results, and one yielded a marginally significant effect in the opposite direction. We consider several possible reasons for the discrepancies (e.g., selective reporting, data anomalies, time-related or context-related moderators) and discuss their theoretical and practical implications. Altogether, our findings imply that the price divisibility effect may not exist; if it does exist, it is not as robust or generalizable as originally claimed. Researchers and marketers should exercise caution before attempting to extend or apply research on the price divisibility effect.

Keywords: multipack, price, quantity, divisibility, replication, open science

A fundamental principle of economics is that, with rare exceptions, higher prices lead to lower demand.

Amanda E. Geiser (ageiser@berkeley.edu) is a PhD candidate in marketing at the University of California, Berkeley, Haas School of Business, 2220 Piedmont Avenue, Berkeley, CA 94720, USA. Leif D. Nelson (leif_nelson@berkeley.edu) is the Ewald T. Grether Professor in Business Administration & Marketing at the University of California, Berkeley, Haas School of Business, 2220 Piedmont Avenue, Berkeley, CA 94720, USA. Address all correspondence to Amanda E. Geiser. We are grateful to members of the Judgment and Decision Making Lab at Berkeley Haas for their helpful feedback on this work. Supplementary materials are included in the [web appendix](#) accompanying the online version of this article.

Editor: Oleg Urminsky

Associate Editor: Bart de Langhe

Advance Access publication December 8, 2025

However, researchers have identified some cases in which consumers are *more* willing to purchase higher-priced products. Here, we investigate the evidence behind a particularly striking example of this phenomenon documented in a recent *Journal of Consumer Research (JCR)* article: the *price divisibility effect* (Park, Kwon, and Bagchi 2024).

The price divisibility effect was demonstrated in the context of multipack purchases. A multipack is a bundle of identical products (e.g., a six-pack of casaba melons). Ostensibly, a multipack is just another product, and therefore, like most products, it should be less appealing to consumers the more expensive it is. However, Park et al. hypothesized and found evidence for a peculiar deviation from this expectation: Because consumers like it when the price of a multipack is divisible by the number of units it contains, they will sometimes prefer a multipack with a higher divisible price over an identical multipack with a lower, but non-divisible, price. For instance, the six-pack of casaba melons might be more appealing if it is priced at

© The Author(s) 2025. Published by Oxford University Press on behalf of Journal of Consumer Research, Inc.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited. • Vol. 00 • 2026
<https://doi.org/10.1093/jcr/ucaf067>

\$12 than if it is priced at \$11, because \$12 is divisible by six. Although consumers generally like to keep their money, sometimes it appears as though they like divisible prices enough to spend more of it.

Park et al. theorized that consumers prefer divisible prices because divisible prices make it easier to calculate per-unit prices, thus shifting consumers' attention from the multipack to the individual units. The more people focus on unit consumption, the more quickly they expect to consume each unit, and the easier it is to justify purchasing a multipack. As a result, the original authors argue, consumers are sometimes even more willing to purchase a multipack when it has a higher divisible price than when it has a lower non-divisible price.

The original article presented a substantial accumulation of evidence across 15 studies (seven in the main text and eight in the web appendix). The original studies demonstrated the price divisibility effect across a variety of product contexts, from toothbrushes to salted cashews; tested several theory-derived mediators (e.g., justifiability) and moderators (e.g., availability of unit prices); and ruled out alternative explanations (e.g., a preference for round-number prices). One study (study 1a) was preregistered, and the evidential value of the full package of studies was assessed using a *p*-curve analysis. As reported in the original article's web appendix, the *p*-curve confirmed that the studies' key effects contained evidential value and estimated their collective statistical power to be 77%. If the original findings are as reliable as they seem, they suggest we should update our understanding of microeconomic theory and encourage marketers to incorporate the price divisibility effect into their toolkits.

Given the theoretical and practical importance of the price divisibility effect, it is worthwhile to test whether the original findings replicate. In order for a scientific claim to be true—let alone have any practical relevance—it must be replicable. Yet large-scale replication efforts have revealed that many published findings have less evidential value than originally claimed (Camerer et al. 2018; O'Donnell et al. 2021; Open Science Collaboration 2015). Conducting replications can help determine not only whether a finding is true but also whether it generalizes beyond a particular sample or a particular point in time (Urminsky and Dietvorst 2024). For instance, if consumers were especially sensitive to price divisibility around the start of the COVID-19 pandemic (when Park et al.'s studies were conducted), then the original findings may not generalize to other points in time, which could pose a challenge to the original theory.

We therefore set out to examine the robustness of the price divisibility effect by conducting preregistered replications of three key studies from the original article (studies 1a, 1b, and 4). In the original study 1a, participants reported being more likely to purchase a four-pack of body wash when it had the divisible price of \$16 than when it

had the lower, but non-divisible, price of \$15.30. The original study 1b conceptually replicated this effect in a new context in which the non-divisible price was a round number, but the divisible price was not. Specifically, participants in this study reported that they would be more likely to purchase an 11-bag multipack of salted cashews if it were priced at \$11 than if it were priced at \$10. Finally, the original study 4 tested a moderator of the price divisibility effect: whether consumers are told how quickly they can expect to consume each unit. Participants reported that they would be more likely to purchase an eight-pack of toothbrushes at the divisible price of \$16 than at the non-divisible price of \$15.41, but this effect disappeared when they were told that each toothbrush would last for two months.

We chose to replicate this particular set of studies for two reasons. First, all three studies were easy to replicate with nearly perfect fidelity. We used the exact same manipulations, stimuli, measures, and analyses as the original studies, which meant that the only differences between our replications and the original studies were theoretically irrelevant factors (e.g., samples, timing). Our replications thus constitute study-design replications (Urminsky and Dietvorst 2024), which allow us to test whether the original findings are robust to factors that should not influence the results. Successful study-design replications can increase the field's confidence in a particular finding; unsuccessful replications might suggest that a finding is not as robust or generalizable as originally claimed, perhaps posing a challenge to the original theory. The second reason we sought to replicate this particular set of studies is that one of them was originally preregistered, and all three original studies yielded strong evidence. For example, the *p*-values for these studies' key results were, respectively, .001, .002, and .005. *P*-curve estimates the combined statistical power of these three studies to be 80%. In other words, if the original studies were replicated many times over, on average, they would be expected to yield statistically significant results 80% of the time. With even larger sample sizes, the overall rate of replication should move closer to 100%.

OVERVIEW OF STUDIES

We conducted four replications in total. Studies 1–3 (reported in the main text) are replications of studies 1a, 1b, and 4, respectively, from the original article. Appendix study A (reported in the web appendix) is an additional replication and extension of study 1a with two additional non-divisible price conditions, which allowed us to examine whether fluctuations in price sensitivity might have contributed to variation in the size of the price divisibility effect.

To evaluate our replication results, we used two different approaches. First, we asked whether the effects observed in the original studies also emerged in our replications.

Second, we asked whether the effects observed in our replications were large enough to have been detectable with the original studies' sample sizes. Following the small-telescopes procedure outlined by [Simonsohn \(2015\)](#), we calculated the minimum effect size (in partial r terms, to capture the direction of the effect) that would have given each original study at least 33% power ($r_{33\%}$). Any effect smaller than this would yield nonsignificant results more than twice as often as significant ones, meaning that the original study would have been severely underpowered. We then tested whether the effect size observed in our replication significantly differed from this threshold. If a replication yields an effect that is significantly smaller than what the original study could have reliably detected, then this implies that even if the effect of interest does exist, the original study's sample would not have been large enough to detect it. We also calculated the implied power of the original study to detect an effect in the hypothesized direction using a two-tailed statistical test. For each replication, we report one power estimate that assumes the true effect size is at the upper boundary of our 95% confidence interval (CI), and another that assumes the true effect size is equal to the effect size observed in our replication.

Across all four of our replications, we found no support for the hypothesis that consumers prefer multipacks with divisible (vs. non-divisible) prices. In our replications of studies 1a and 4 from the original article (studies 1, 3, and appendix study A), purchase likelihood did not differ for multipacks with higher divisible prices versus those with lower non-divisible prices, regardless of how large the gap was between the divisible and non-divisible prices (appendix study A) or whether participants were told how long each unit would last (study 3). In our replication of study 1b from the original article (study 2), purchase likelihood was somewhat *lower* for a multipack with a higher divisible price than for one with a lower non-divisible price. Even if we assume the true effect sizes are at the upper limit of our 95% CIs, the original studies would have had between 2.6% and 35.6% power to detect those effects. If we assume the effect sizes observed in our replications are correct, the original studies would have only had between 0.07% and 5.4% power. [Figure 1](#) displays the results of each original study alongside our main set of replication results, and [figure 2](#) depicts effect-size estimates for each study.

All studies were preregistered on AsPredicted. In line with the recommendations of [Simonsohn \(2015\)](#), we aimed to recruit approximately 2.5 times as many participants per condition as the original studies. Participants in studies 1–3 were recruited from Prolific, and participants in appendix study A were recruited from Amazon Mechanical Turk (MTurk) via CloudResearch. The target sample sizes for all studies were specified in advance in our preregistrations, as were all exclusion criteria, manipulations, measures, and analyses. Our preregistrations, materials, data, and code are

available on ResearchBox (https://researchbox.org/2281&PEER_REVIEW_passcode=NRSQBL).

STUDY 1: REPLICATION OF PARK ET AL.'S STUDY 1A

Study 1a in the original article demonstrated the price divisibility effect in consumers' preferences for a body-wash multipack. Specifically, participants ($N=197$) reported that they would be more likely to purchase a four-pack of body wash if it had the divisible price of \$16 than if it had the lower, but non-divisible, price of \$15.30. Our first study sought to replicate this pattern. We chose to replicate this study in part because it was the only preregistered study reported in the original article, which meant we could more easily ensure that the design, procedure, and analysis plan of our replication exactly matched the original.

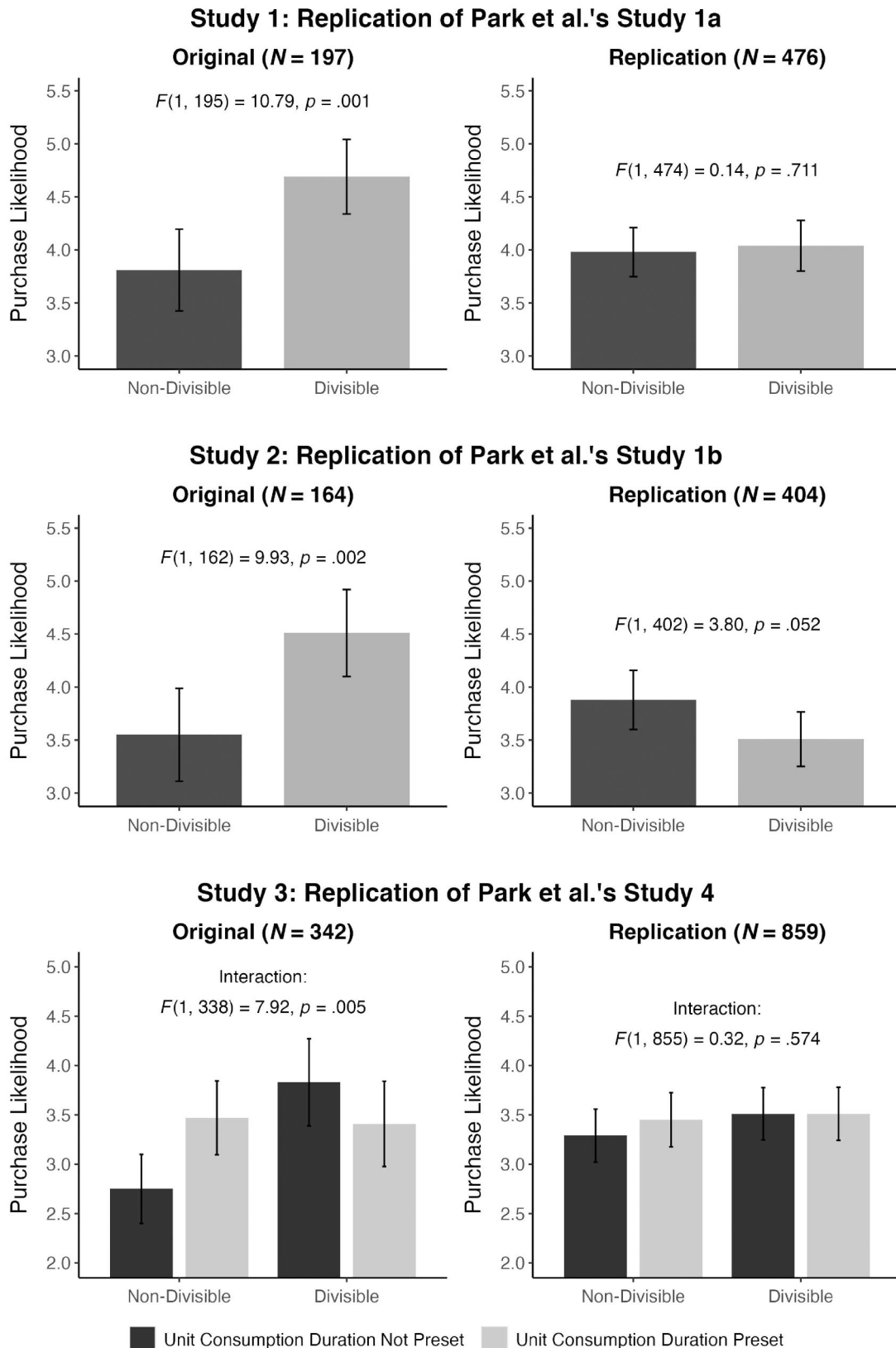
Method

Participants and Design. We requested 750 participants from Prolific and received 743 complete submissions. As preregistered (<https://aspredicted.org/hr9t-38tc.pdf>), we excluded participants who failed an attention check at the start of the survey ($n=24$), using the same procedure as the original study. Our final sample included 719 participants ($M_{age}=41.33$ years; gender = 48.54% men, 49.79% women, 1.67% other identity). Participants were randomly assigned to either a divisible \$16 price condition, a non-divisible \$15.3 price condition (one decimal digit), or a non-divisible \$15.30 price condition (two decimal digits). The first two conditions ($n=476$) are identical to those in the original study, and the third is a new condition that we added so that we could ensure that none of the results hinge on showing the non-divisible price with only a single decimal digit.

Procedure. Participants saw a picture of a four-pack body-wash multipack priced at either "\$16" (divisible price condition), "\$15.3" (non-divisible price with one decimal digit condition), or "\$15.30" (non-divisible price with two decimal digits condition). They indicated how likely they would be to purchase the multipack: "How likely are you to purchase this item?" (1 = "very unlikely," 7 = "very likely"). Next, they completed three items that were intended to measure price divisibility: "It was easy for me to calculate how much each pack of body wash was priced at," "The price of each pack of body wash easily came to my mind," and "The price of each pack of body wash was clear to me" (1 = "strongly disagree," 7 = "strongly agree"). Responses to these three items were averaged to create a price divisibility composite ($\alpha=0.93$). Finally, participants reported their age and gender.

FIGURE 1

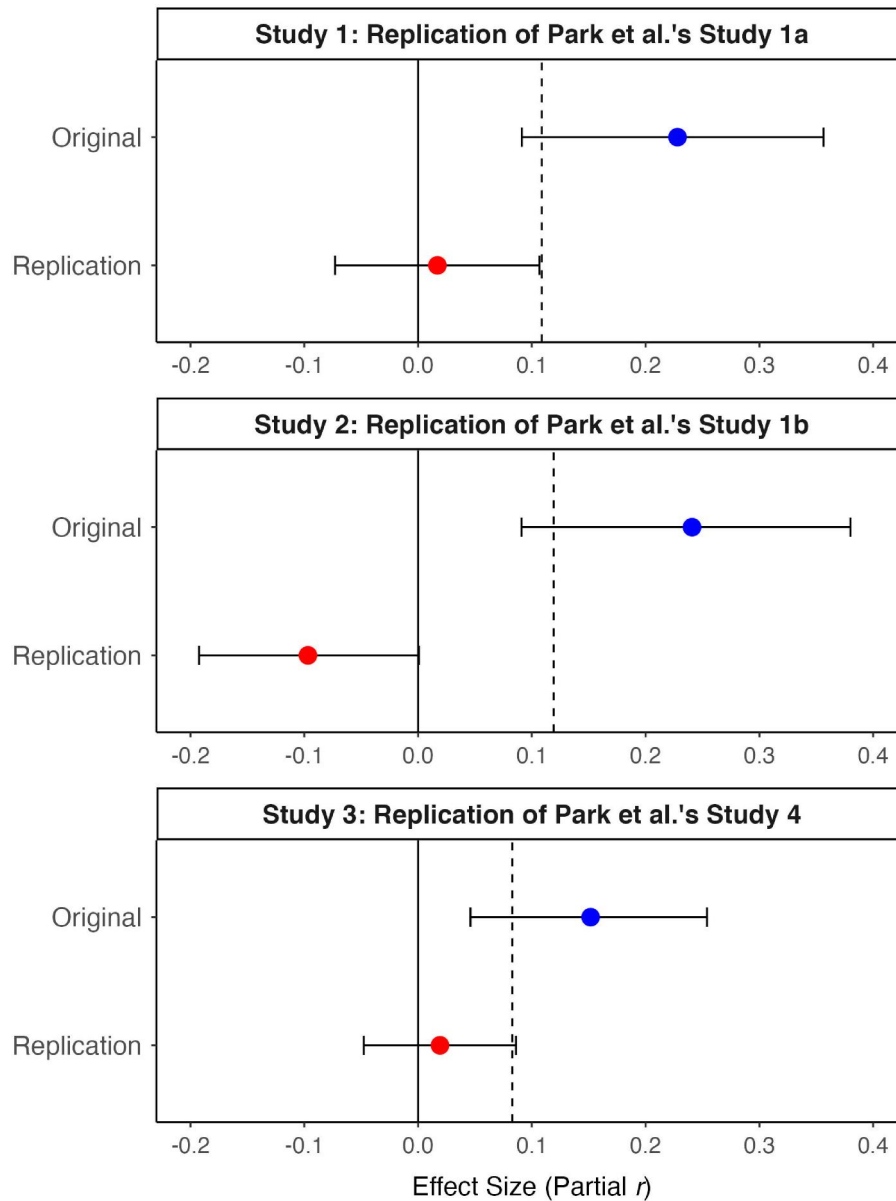
PURCHASE LIKELIHOOD BY CONDITION IN EACH ORIGINAL STUDY (LEFT) AND REPLICATION (RIGHT)



NOTE.—Error bars represent 95% confidence intervals.

FIGURE 2

EFFECT SIZE IN EACH ORIGINAL AND REPLICATION STUDY



NOTE.—Error bars represent 95% confidence intervals. Positive effect sizes indicate effects consistent with the direction of the original findings. The dashed line indicates the effect size ($r_{33\%}$) that would have given the original study 33% power.

Results

Manipulation Check. Participants perceived the price of the body-wash multipack as more easily divisible when it was priced at \$16 ($M = 6.07$, $SD = 1.24$) than when it was priced at \$15.30, regardless of whether the non-divisible price was shown with one decimal digit

($M = 5.15$, $SD = 1.48$; $F(1, 474) = 53.96$, $p < .001$, partial $\eta^2 = 0.102$) or two decimal digits ($M = 5.13$, $SD = 1.66$; $F(1, 482) = 50.13$, $p < .001$, partial $\eta^2 = 0.094$). As shown in table 1, this pattern of means is nearly identical to the pattern observed in the original study. This suggests that the price divisibility manipulation operated as intended, but

TABLE 1
SUMMARY OF RESULTS BY CONDITION IN PARK ET AL.'S STUDY 1A AND OUR REPLICATION (STUDY 1)

Study	Dependent variable	Non-divisible \$15.3 price mean (SD)	Non-divisible \$15.30 price mean (SD)	Divisible \$16 price mean (SD)	Effect sizes (<i>r</i>)	
					\$15.3 versus \$16	\$15.30 versus \$16
Park et al.'s study 1a	Purchase likelihood	3.81 (1.95)	—	4.69 (1.78)	0.228	—
	Price divisibility composite	5.11 (1.63)	—	6.06 (1.27)	0.311	—
Replication	Purchase likelihood	3.98 (1.81)	3.97 (1.85)	4.04 (1.89)	0.017	0.019
	Price divisibility composite	5.15 (1.48)	5.13 (1.66)	6.07 (1.24)	0.320	0.307

NOTE.—This table shows means and SDs of each measure by condition and effect sizes for the difference between the divisible and non-divisible price conditions. Positive effect sizes indicate that responses were greater, on average, in the divisible price condition than in the non-divisible price condition.

also that the manipulation had a similar effect in our replication as it did in the original study.

Purchase Likelihood. Whereas participants in the original study were significantly more likely to purchase the body-wash multipack when it was priced at \$16 than when it was priced at \$15.30, participants in our replication were no more likely to purchase the body-wash multipack when it was priced at \$16 ($M = 4.04$, $SD = 1.89$) than when it was priced at \$15.30, regardless of whether the latter price was shown with one decimal digit ($M = 3.98$, $SD = 1.81$; $F(1, 474) = 0.14$, $p = .711$, partial $\eta^2 = 0.0003$) or two decimal digits ($M = 3.97$, $SD = 1.85$; $F(1, 482) = 0.17$, $p = .679$, partial $\eta^2 = 0.0004$). In other words, neither effect was detectably different from zero.

Effect-Size Comparison. The original study had 197 participants, giving it 33% power to detect $r = 0.109$. In other words, for the original study to detect an effect at least a third of the time, the true effect size would need to be at least $r = 0.109$. In our replication, we observed an effect size of $r = 0.017$ (95% CI $[-0.073, 0.107]$) for the comparison of the \$16 divisible price to the \$15.3 non-divisible price (as shown in figure 2) and an effect size of $r = 0.019$ (95% CI $[-0.070, 0.108]$) for the comparison of the \$16 divisible price to the \$15.30 non-divisible price. Both of these effects are significantly smaller than $r_{33\%}$. If we assume the true effect size is at the upper limit of our 95% CI ($r = 0.107$) for the comparison of \$16 to \$15.3, the original study would have had 32.1% power to detect it, and if we assume the true effect size is equal to our best estimate of that effect ($r = 0.017$), the original study would have had only 4.2% power to detect it. Likewise, if we assume the true effect size is at the upper limit of our 95% CI ($r = 0.108$) for the comparison of \$16 to \$15.30, the original study would have had only 32.6% power to detect it, and if we assume the true effect size is equal to our best estimate of that effect ($r = 0.019$), the original study would have had 4.5% power to detect it. Altogether, our results suggest that if the effect of interest exists, the original study's sample size would not have been large enough to detect it.

Discussion

In sum, we found no support for Park et al.'s claim that consumers are more likely to purchase a four-pack of body wash when it is priced at \$16 than when it is priced at \$15.30. Neither of the effects observed in our replication was detectably different from zero, and both effects were significantly smaller than the smallest effect that the original study's sample size could have reliably detected. Our results did not hinge on whether the non-divisible price was shown with only a single decimal digit (as in the original study) or with two decimal digits (as is standard in the marketplace).

One potential concern is that the prices used in the original study were perceived differently by participants in our replication. For instance, perhaps consumers were more price-sensitive in 2023 (when we conducted our replication) than in 2021 (when the original study was conducted), in which case the gap between the divisible and non-divisible prices may have seemed larger to our replication participants, washing out the benefits of divisible prices. If so, then the price divisibility effect may reemerge when the non-divisible and divisible prices are closer together. Appendix study A (reported in the web appendix) tested for this possibility by including two additional non-divisible price conditions (\$15.60 and \$15.90) in between the two original prices (\$15.30 and \$16). As in study 1, the price divisibility effect did not emerge when comparing the divisible \$16 price ($M = 4.12$, $SD = 1.94$) to the non-divisible \$15.30 price used in the original study ($M = 4.34$, $SD = 1.77$; $t(985) = 1.36$, $p = .386$). However, the effect also did not emerge when the non-divisible price was \$15.60 ($M = 4.42$, $SD = 1.87$; $t(985) = 1.79$, $p = .181$), nor when the non-divisible price was \$15.90 ($M = 4.08$, $SD = 1.86$; $t(985) = -0.24$, $p = .991$). In two out of three cases, participants were directionally *less* likely to purchase the multipack with the higher divisible price than the one with the lower non-divisible price. This suggests that it is unlikely that we failed to detect a price divisibility effect in study 1 because our participants were more price-sensitive than participants in the original study.

STUDY 2: REPLICATION OF PARK ET AL.'S STUDY 1B

Our second study was a replication of study 1b in the original article, which demonstrated the price divisibility effect in a new context in which the non-divisible price was a round number, whereas the divisible price was not. If the price divisibility effect is driven by a preference for divisibility itself, as opposed to a preference for round numbers, then this effect should emerge even when the divisible price is not a round number. Consistent with this idea, participants in the original study ($N = 164$) reported that they would be more likely to purchase an 11-bag multipack of salted cashews when it was priced at \$11 than when it was priced at \$10.

Method

Participants and Design. We requested 415 participants from Prolific and received 416 complete submissions. As preregistered (<https://aspredicted.org/9kmm-kmrb.pdf>), we excluded participants who failed an attention check at the start of the survey ($n = 12$), using the same procedure as the original study. Our final sample included 404 participants ($M_{\text{age}} = 41.28$ years, gender = 50.74% men, 48.51% women, 0.74% another identity). Participants were randomly assigned to either a divisible price (\$11) or non-divisible price (\$10) condition.

Procedure. Participants saw a picture of an 11-bag salted cashew multipack priced at either \$11 (divisible price condition) or \$10 (non-divisible price condition). They indicated how likely they would be to purchase the multipack: "How likely are you to purchase this multipack?" (1 = "very unlikely," 7 = "very likely").

Although the original study did not include a manipulation check, we adapted the three items used in study 1a to assess perceptions of price divisibility: "It was easy for me to calculate how much each pack of cashews was priced at," "The price of each pack of cashews easily came to my mind," and "The price of each pack of cashews was clear to me" (1 = "strongly disagree," 7 = "strongly agree"). Responses to these three items were averaged to create a price divisibility composite ($\alpha = 0.92$). Finally, participants reported their age and gender.

Results

Manipulation Check. Participants perceived the price of the 11-bag cashew multipack as more easily divisible when it was \$11 ($M = 5.40$, $SD = 1.71$) than when it was \$10 ($M = 5.10$, $SD = 1.56$), although this difference was only marginally significant ($F(1, 402) = 3.38$, $p = .067$, partial $\eta^2 = 0.008$). Given that the original study did not measure perceptions of divisibility, we cannot determine whether the manipulation had a larger or smaller effect in

our replication than in the original study. However, our results provided at least directional support for the assumption that the \$11 price was perceived as more divisible than the \$10 price.

Purchase Likelihood. Whereas participants in the original study reported being significantly more likely to purchase the cashew multipack when it was priced at \$11 than when it was priced at \$10 (table 2), our replication yielded a marginally significant effect in the opposite direction. That is, our participants reported that they would be somewhat *less* likely to purchase the \$11 cashew multipack ($M = 3.51$, $SD = 1.88$) than the \$10 cashew multipack ($M = 3.88$, $SD = 2.00$; $F(1, 402) = 3.80$, $p = .052$, partial $\eta^2 = 0.009$).

Effect-Size Comparison. The original study had 164 participants, giving it 33% power to detect $r = 0.119$. In other words, for the original study to detect an effect at least a third of the time, the true effect size would need to be at least $r = 0.119$. As shown in figure 2, our replication yielded an effect size of $r = -0.097$ (95% CI [-0.193, 0.001]), which is significantly below this threshold. The upper end of our CI allows for the possibility of a positive effect, albeit a very small one ($r = 0.001$). If we assume the true effect size is at the upper boundary of our CI, the original study would have had 2.6% power to detect it. If we assume the true effect size is equal to our best estimate ($r = -0.097$), the original study would have had only 0.07% power to detect an effect in the hypothesized direction. Altogether, our results imply that if the effect of interest exists, the original study's sample size would not have been large enough to detect it.

Discussion

In sum, we found no support for Park et al.'s claim that consumers are more likely to purchase an 11-bag multipack of salted cashews when it is priced at \$11 than when it is priced at \$10. Although participants in our replication perceived the \$11 price as somewhat *more* divisible than the \$10 price, they were marginally significantly *less* likely to purchase the \$11 multipack than the \$10 multipack. Our observed effect size differed significantly from the smallest effect that could have been reliably detected with the original study's sample size.

STUDY 3: REPLICATION OF PARK ET AL.'S STUDY 4

In our final study, we set out to replicate study 4 from the original article, which tested whether the price divisibility effect is moderated by the presence of unit-consumption-duration information. If consumers prefer divisible prices because they expect to consume each unit in a multipack more quickly when its price is divisible by

TABLE 2
SUMMARY OF RESULTS BY CONDITION IN PARK ET AL.'S STUDY 1B AND OUR REPLICATION (STUDY 2)

Study	Dependent variable	Non-divisible \$10 price mean (SD)	Divisible \$11 price mean (SD)	Effect size (<i>r</i>)
Park et al.'s study 1b Replication	Purchase likelihood	3.55 (2.02)	4.51 (1.89)	0.241
	Purchase likelihood	3.88 (2.00)	3.51 (1.88)	-0.097
	Price divisibility composite	5.10 (1.56)	5.40 (1.71)	0.091

NOTE.—This table shows means and SDs of each measure by condition and effect sizes for the difference between the non-divisible and divisible price conditions. Positive effect sizes indicate that responses were greater, on average, in the divisible price condition than in the non-divisible price condition.

the number of units it contains—as the original authors theorize—then the effect may disappear when unit-consumption-duration information is provided. Consistent with this hypothesis, participants in the original study ($N = 342$) reported that they would be more likely to purchase an eight-pack toothbrush multipack priced at \$16 than one priced at \$15.41, but only if they did not receive information about unit-consumption duration. Among those who did receive unit-consumption-duration information, purchase likelihood was no greater for the \$16 multipack than for the \$15.41 multipack.

Method

Participants and Design. We requested 868 participants from Prolific and received 865 complete submissions. As preregistered (<https://aspredicted.org/ssv4-7ywy.pdf>), we excluded participants who failed an attention check at the start of the survey ($n = 6$), using the same procedure as the original study. Our final sample included 859 participants ($M_{\text{age}} = 41.00$ years, gender = 49.94% men, 50.06% women). Participants were randomly assigned to one of four conditions in a 2 (price divisibility: divisible price vs. non-divisible price) by 2 (unit-consumption duration: preset vs. not preset) between-subjects design.

Procedure. Participants saw a picture of an eight-pack charcoal toothbrush multipack priced at either \$16 (divisible price condition) or \$15.41 (non-divisible price condition). Those in the unit-consumption-duration-preset condition read that “The floss-tip bristles last for 2 months after being exposed to water. The toothbrush should be replaced every 2 months.” Meanwhile, those in the unit-consumption-duration-not-preset condition were not informed about the expected consumption duration of each toothbrush and simply read, “Floss-tip bristles.” They indicated how likely they would be to purchase the multipack: “How likely are you to purchase this toothbrush multipack?” (1 = “very unlikely,” 7 = “very likely”).

Next, participants responded to three items that assessed the justifiability of purchasing the multipack: “How justifiable is it to purchase multiple units of toothbrush?” (1 = “not justifiable at all,” 7 = “very justifiable”), “How easy to defend is it to purchase multiple units of toothbrush?” (1

= “not easy to defend at all,” 7 = “very easy to defend”), and “How logical is it to purchase multiple units of toothbrush?” (1 = “not logical at all,” 7 = “very logical”). Responses to these three items were averaged to create a justifiability composite ($\alpha = 0.94$). They next completed a measure of expected quickness of unit consumption: “I think each toothbrush will last _____.” (1 = “very short,” 7 = “very long”). As in the original study, we reverse-coded responses to this item so that higher values would indicate greater quickness. Finally, participants reported their age and gender.

Results

Purchase Likelihood. Table 3 shows the results by condition for both the original study and our replication. Park et al. found that purchase likelihood for an eight-pack toothbrush multipack with a divisible (\$16) versus non-divisible (\$15.41) price hinged on whether unit-consumption-duration information was provided. In our replication, the main effect of price divisibility was not significant ($F(1, 855) = 1.05, p = .305$, partial $\eta^2 = 0.001$), indicating that participants were no more likely to purchase the \$16 toothbrush multipack ($M = 3.51, SD = 1.99$) than the \$15.41 toothbrush multipack ($M = 3.37, SD = 2.03$). The main effect of unit-consumption-duration information was also not significant ($F(1, 855) = 0.34, p = .561$, partial $\eta^2 = 0.0004$), indicating that participants were no more likely to purchase the multipack when they were told that each toothbrush would last for two months ($M = 3.48, SD = 2.02$) than when they were not told how long each toothbrush would last ($M = 3.40, SD = 2.00$).

The original study's central finding was that the effect of price divisibility depended on whether unit-consumption-duration information was provided, such that participants preferred the \$16 multipack over the \$15.41 multipack only when they were not told how long each toothbrush would last. By contrast, in our replication, the effect of price divisibility did not depend on whether unit-consumption-duration information was provided ($F(1, 855) = 0.32, p = .574$, partial $\eta^2 = 0.0004$). The price divisibility effect did not emerge when participants were told that each toothbrush would last for two months ($M_{\text{divisible}} =$

TABLE 3

SUMMARY OF RESULTS BY CONDITION IN PARK ET AL.'S STUDY 4 AND OUR REPLICATION (STUDY 3)

Study	Dependent variable	Non-divisible price mean (SD)	Divisible price mean (SD)	Interaction effect size (partial <i>r</i>)
Park et al.'s study 4 (N = 342)	Purchase likelihood	Duration not preset: 2.75 (1.65)	Duration not preset: 3.83 (2.08)	0.152
		Duration preset: 3.47 (1.77)	Duration preset: 3.41 (2.04)	
	Expected quickness of unit consumption	Duration not preset: 3.58 (1.68)	Duration not preset: 4.25 (1.60)	0.100
		Duration preset: 4.82 (1.53)	Duration preset: 4.87 (1.30)	
	Justifiability	Duration not preset: 4.39 (1.94)	Duration not preset: 5.12 (1.76)	0.148
		Duration preset: 5.12 (1.79)	Duration preset: 4.74 (1.90)	
Replication (N = 859)	Purchase likelihood	Duration not preset: 3.29 (2.01)	Duration not preset: 3.51 (1.99)	0.019
		Duration preset: 3.45 (2.05)	Duration preset: 3.51 (2.00)	
	Expected quickness of unit consumption	Duration not preset: 3.95 (1.41)	Duration not preset: 3.82 (1.26)	0.010
		Duration preset: 4.64 (1.48)	Duration preset: 4.56 (1.51)	
	Justifiability	Duration not preset: 5.05 (1.76)	Duration not preset: 5.27 (1.67)	0.042
		Duration preset: 5.40 (1.62)	Duration preset: 5.34 (1.59)	

NOTE.—This table shows means and SDs of each measure by condition and effect sizes for the two-way interaction between price divisibility and the presence of unit-consumption-duration information. Positive effect sizes indicate effects consistent with the direction of the original findings.

3.51, SD = 2.00 vs. $M_{\text{non-divisible}} = 3.45$, SD = 2.05; $F(1, 425) = 0.10$, $p = .747$, partial $\eta^2 = 0.0002$), nor did it emerge when they were not told how long each toothbrush would last ($M_{\text{divisible}} = 3.51$, SD = 1.99 vs. $M_{\text{non-divisible}} = 3.29$, SD = 2.01; $F(1, 430) = 1.28$, $p = .259$, partial $\eta^2 = 0.003$).

Effect-Size Comparison. The original study had 342 participants, giving it 33% power to detect an $r = 0.083$ interaction effect. In other words, the true effect size would need to be at least $r = 0.083$ to be detected with the original study's sample size at least a third of the time. As shown in figure 2, our replication yielded an effect size of $r = 0.019$ (95% CI [-0.048, 0.086]), which is not detectably different from the smallest effect that would have given the original study at least 33% power. However, even if we assume the true effect size is at the upper boundary of our CI ($r = 0.086$), the original study would still have had only 35.6% power to detect it. If we assume the true effect size is equal to our best estimate ($r = 0.019$), the original study would have had 5.4% power to detect it. In sum, our results allow for the possibility that the effect of interest is very small yet just large enough to have been detectable with the original study's sample size. However, our results are most consistent with the absence of any effect.

Expected Quickness of Unit Consumption. As in the original study, participants expected to consume each unit

more quickly when they were told that each toothbrush would last for two months ($M = 4.60$, SD = 1.50) than when they were not told how long each toothbrush would last ($M = 3.89$, SD = 1.34; $F(1, 855) = 54.50$, $p < .001$, partial $\eta^2 = 0.060$). However, unlike in the original study, they did not expect to consume the multipack any more quickly when it was priced at \$16 ($M = 4.19$, SD = 1.44) than when it was priced at \$15.41 ($M = 4.30$, SD = 1.49; $F(1, 855) = 1.21$, $p = .272$, partial $\eta^2 = 0.001$), nor did this effect depend on whether information about unit-consumption duration was provided ($F(1, 855) = 0.09$, $p = .767$, partial $\eta^2 = 0.0001$). This means that participants did not expect to consume each unit more quickly when the multipack had a divisible price than when it had a non-divisible price, regardless of whether they were told how long each toothbrush would last.

Justifiability. Consistent with the original study, participants did not see purchasing the \$16 multipack as more justifiable ($M = 5.30$, SD = 1.63) than purchasing the \$15.41 multipack ($M = 5.22$, SD = 1.70; $F(1, 855) = 0.54$, $p = .465$, partial $\eta^2 = 0.0006$). Whereas justifiability was not affected by the presence of unit-consumption-duration information in the original study, participants in our replication saw purchasing the toothbrush multipack as somewhat more justifiable if they were told that each toothbrush would last for two months ($M = 5.37$, SD = 1.60) than if they were not ($M = 5.16$, SD = 1.72; $F(1, 855) = 3.37$, $p =$

.067, partial $\eta^2 = 0.004$). Moreover, unlike the original study, we did not detect an interaction between price divisibility and unit-consumption-duration information ($F(1, 855) = 1.49, p = .223$, partial $\eta^2 = 0.002$). In other words, although participants thought it would be more justifiable to purchase the toothbrush multipack when told how long each toothbrush would last, their perceptions of justifiability did not depend on whether the multipack had a divisible or non-divisible price.

Discussion

Altogether, our findings do not support the original theory. Park et al.'s study 4 found that participants preferred an eight-pack of toothbrushes when it was priced at \$16 than when it was priced at \$15.41, but only if they were not told how long each toothbrush would last. By contrast, in our replication, a price divisibility effect did not emerge regardless of whether information about unit-consumption-duration was provided. The focal interaction effect observed in the original study was not detectably different from zero in our replication. Nevertheless, the effect observed in our replication was also not detectably different from the smallest effect that the original study's sample size could have detected. This means we cannot rule out the possibility that the effect of interest is large enough to have been detectable by the original study.

The additional measures of expected quickness of unit consumption and justifiability also failed to yield support for the original theory. Park et al. theorized that the price divisibility effect emerges because divisible prices make unit consumption more salient, thus leading consumers to expect that each unit will be consumed more quickly and making it seem more justifiable to purchase a multipack. Although participants in our replication expected to consume each unit more quickly and thought purchasing the multipack was more justifiable when they were told how long each unit would last (as in the original study), they were not influenced at all by the divisibility of the multipack's price.

GENERAL DISCUSSION

Research on the price divisibility effect suggested that consumers prefer multipacks with higher divisible prices over those with lower non-divisible prices. However, across four preregistered replications, we found no evidence for this phenomenon. In fact, one of our replications yielded a marginally significant effect in the opposite direction of the original (i.e., greater purchase likelihood for a lower non-divisible price than for a higher divisible price). Our findings suggest that if the effects of interest do exist, two out of the three original studies' samples were too small to detect them.

Why did the price divisibility effect emerge in the original studies yet not in our replications? There are several possible explanations, but first it is worthwhile to acknowledge what is probably not a good explanation: bad luck. It is unlikely that we failed to replicate the price divisibility effect entirely due to chance. All three of the original studies strongly rejected the null hypothesis in favor of the alternative, and none of their key results were perilously close to the $p = .05$ threshold. Moreover, our replications all had much larger sample sizes, affording them increased statistical power and allowing us to detect effects even smaller than those in the original studies. If chance cannot adequately explain why our replications failed, then what can?

One possible explanation is that the original findings were selectively reported. Because nonsignificant findings are less likely to be published (Sterling 1959), authors have learned to stop submitting them. This means that the published literature necessarily overestimates the prevalence and magnitude of true effects and that some published findings are merely false positives. If one conducts enough studies and/or analyses, they will eventually obtain significant results even if the effect of interest does not exist (Simmons, Nelson, and Simonsohn 2011). However, this explanation does not seem likely, at least on its own. After all, the original studies yielded results consistent with strong evidential value, and the three studies that we chose to replicate all had p -values below .01. Moreover, one of the original studies was preregistered, thus precluding many possible researcher degrees of freedom.

Another possibility is that there were features of the original materials that did not appear in our replication materials. For example, small differences in the exact flow or formatting of the surveys may have inadvertently favored divisible (vs. non-divisible) prices in the original studies or counteracted the advantage of divisible (vs. non-divisible) prices in our replications. For our replications of studies 1a and 1b, we based our materials on the ones reported in the original article's web appendix, which means that they may have differed from the original studies in minor ways; however, in our replication of study 4, we circumvented this issue by using the original study's QSF file. It is therefore unlikely that our replications failed to yield support for the price divisibility effect solely due to differences in materials.

It is also possible that there are issues with or anomalies in the underlying data. Perhaps there were errors in the coding or reporting of the original or replication data, or perhaps there is some anomalous feature of the original data that cannot be assessed via summary statistics alone (e.g., a systematic response bias that only affected the original studies). Of course, the only way to identify such issues is to closely examine both sets of data. Because the original studies' data are not publicly available, we were not able to examine them directly. However, our replication data are

posted publicly on ResearchBox. Although we are not aware of any issues and/or anomalies in our data, we encourage interested readers to check for themselves.

Another possible reason for the discrepancies between our results and the results of the original studies is that some essential component of the original theory was missing in our replications or that the original theory failed to specify all of the conditions that are necessary for the price divisibility effect to emerge. In concept, our replications could have failed to detect a price divisibility effect either because divisible prices were not perceived as more divisible than non-divisible prices or because greater perceived divisibility did not translate into greater purchase likelihood. Given that three out of our four replications measured perceptions of divisibility and found that divisible prices were indeed perceived as more divisible than non-divisible prices, the first possibility seems unlikely. However, it is possible that there is an important moderator of the relationship between perceived divisibility and purchase likelihood that operated differently in our replications compared to the original studies.

Perhaps the most obvious sources of variation between the original studies and our replications are the participants who completed them and the time at which they were conducted. If the original findings were specific to a particular sample or a particular point in time, then this might explain why our replications yielded different results (Urminsky and Dietvorst 2024). Participants in our replications were recruited from two different online platforms (Prolific and CloudResearch) that are similar to the one used for the original studies (MTurk) and have at least partially overlapping participant pools. This means it is unlikely that our replication samples systematically differed from the original studies' samples in ways that would have yielded different results. A more plausible explanation is that our results differed because the relationship between perceived divisibility and purchase likelihood changed over time, given that the original studies were conducted in 2021, whereas our replications were conducted in 2023 and 2025.

At the time when the original studies were conducted, the COVID-19 pandemic may have led consumers to adopt unusual purchasing habits that might have made them more susceptible to the effects of divisible prices. Although the original theory did not include any pandemic-specific moderators, it is possible that the original studies were conducted at a time when consumers were especially motivated to stockpile products and thus especially interested in purchasing multipacks. If greater baseline interest in stockpiling makes multipacks with divisible prices seem especially appealing, then this might explain why the price divisibility effect did not emerge a few years later when we conducted our replications. In appendix study A, we explored whether the price divisibility effect is moderated by an individual's tendency to stockpile products. If changes in consumers' stockpiling habits over time explain

why our replication results differed from the original results, then those who chronically stockpile products should be especially susceptible to the effects of divisible prices. Although those who stockpile products more frequently reported greater purchase likelihood overall, we did not find any consistent relationship between an individual's tendency to stockpile and the size of the price divisibility effect. This suggests that it is unlikely that the price divisibility effect disappeared in our replications due to temporal shifts in consumers' stockpiling habits.

Another reason the price divisibility effect could have changed over time is if the prices used in the original studies were perceived differently when we conducted our replications. If the original prices simply seemed higher or lower to participants in our replications across the board, then this would not necessarily explain why we observed smaller *differences* between divisible and non-divisible prices. However, such discrepancies could be explained by fluctuations in price sensitivity. For example, if consumers became more price-sensitive over time, then the objective differences between the original divisible and non-divisible prices may have seemed larger by the time we conducted our replications, perhaps overpowering the psychological benefits of divisibility. To examine this possibility, we tested in appendix study A whether the price divisibility effect reemerges when the divisible and non-divisible prices are closer together. We found that participants were no more likely to purchase the four-pack of body wash from study 1a when it was priced at \$16 than when it was priced at \$15.30 (as in the original study), \$15.60, or \$15.90. This suggests that differences in price sensitivity are unlikely to explain why the price divisibility effect did not emerge in our replications.

Nevertheless, there may still be other important factors that differed between the original studies and our replications that neither we nor the original authors have considered. To fully account for the discrepancies between the original results and our replication results, any proposed moderator should be able to explain not only why divisible prices did not increase purchase likelihood in our replications but also why divisible prices—despite being perceived as more divisible—did not influence expected quickness of unit consumption or the justifiability of purchasing a multipack (which, according to the original theory, serially mediate the price divisibility effect). Unlike in the original study 4, participants in our replication (study 3) did not expect to consume each toothbrush in an eight-pack of toothbrushes any more quickly when the multipack was priced at \$16 rather than \$15.41, nor did they see purchasing the \$16 multipack as any more justifiable. Although there is no particular theoretical reason to expect any of these effects to differ between the original studies and our replications, it is possible that the original theory was merely incomplete and should be updated to accommodate additional moderators.

Altogether, given the large discrepancies between the original results and our replication results, it is unclear whether the price divisibility effect is a real phenomenon. We found no evidence that consumers prefer divisible over non-divisible prices across any of our four replication attempts, despite using the exact same materials and measures as the original studies. We also found no evidence that price divisibility influences other related judgments, such as how quickly consumers expect to consume each unit in a multipack or how justifiable it seems to purchase a multipack. Of course, no single study (or set of studies) can definitively prove that a phenomenon does not exist, and further research would be necessary to determine whether there are any circumstances in which consumers do reliably prefer divisible prices. However, if the price divisibility effect does exist, our results suggest that the conditions under which it occurs, and the range of outcomes that are affected by price divisibility, are far more narrow than originally claimed.

Based on our findings, several of the original article's conclusions may need to be revisited. First, if the price divisibility effect were a real phenomenon, it would pose a serious challenge to standard economic theory by suggesting that price divisibility can counteract or even reverse downward-sloping demand. However, our results did not yield any surprising violations of downward-sloping demand: We did not find any cases in which consumers significantly preferred a more expensive multipack with a divisible price over a less expensive one with a non-divisible price, and in one study, we even found that participants marginally preferred the latter. Our results thus call into question whether price divisibility is powerful enough to counteract the near-universal tendency for consumers to prefer lower prices.

Second, our findings also suggest that marketing managers should exercise caution before incorporating the price divisibility effect into the pricing of multipack products. Park et al. recommend that marketers set multipack prices such that they are divisible by the number of units, either by increasing prices or adjusting the number of units in each multipack. However, our findings suggest that this strategy may not be effective at boosting sales and profits and may even backfire, given that many consumers are likely to be deterred by higher prices. Before implementing divisible pricing in the marketplace, managers should test whether the price divisibility effect emerges in that particular context and with the specific set of prices being considered. The original authors also suggest that marketers may benefit from using divisible prices in other contexts, such as for multiple-unit pricing deals (e.g., “4 for \$8” instead of “4 for \$7”) and single-unit products (e.g., \$12 instead of \$11.43 for a product with 12 servings). Given that we found no benefits of divisible prices even within the context of multipack products, where the phenomenon was originally demonstrated, it may not be wise for marketers

to extend the price divisibility effect to other contexts without further evidence of its effectiveness.

This investigation highlights several ways in which consumer researchers can make it easier for others to replicate, extend, and apply their findings. The aim of science is to identify truths about the world, and truths are those things that can be independently verified, predicted, and replicated. Ideally, those truths should generalize beyond the particular samples, stimuli, manipulations, and measures used to test those claims (Yarkoni 2022). One way that researchers can make it easier for others to directly replicate their findings is by making all of their original data and materials publicly available. Posting data and materials can help future researchers identify potential issues (e.g., confounds, data anomalies) or boundary conditions that might affect the validity or generalizability of the original claims. Researchers can also make it easier for others to build on their findings and test their theories by specifying the conditions under which they expect a particular phenomenon to emerge (e.g., when and for whom). One way to do this is by including a “constraints on generality” section in the general discussion, a practice that some journals already require (Simons, Shoda, and Lindsay 2017).

In sum, research on the price divisibility effect uncovered a serious anomaly in standard economic theory by demonstrating that consumers are sometimes willing to pay more for multipacks with divisible prices. Yet four preregistered replications failed to detect any evidence for this phenomenon, thus calling into question whether it is as robust and generalizable as originally claimed. Although there are many possible reasons why our replication results might have differed from the original results (e.g., selective reporting, data anomalies, time-related or context-related moderators not specified in the original theory), our findings suggest that researchers and marketers should exercise caution before extending research on the price divisibility effect or implementing it in the marketplace.

DATA COLLECTION STATEMENT

The first author collected data for all four studies. Studies 1 and 2 were conducted on Prolific in November 2023, and study 3 was conducted on Prolific in December 2023. Appendix study A, reported in the [web appendix](#), was conducted on CloudResearch in March 2025. The first author analyzed the data, with input from the second author. All data, materials, preregistrations, and code are available on ResearchBox: https://researchbox.org/2281&PEER_REVIEW_passcode=NRSQBL.

REFERENCES

- Camerer, Colin F., Anna Dreber, Felix Holzmeister, et al. (2018), “Evaluating the Replicability of Social Science Experiments

- in Nature and Science Between 2010 and 2015,” *Nature Human Behaviour*, 2 (9), 637–44.
- O’Donnell, Michael, Amelia S. Dev, Stephen Antonoplis, et al. (2021), “Empirical Audit and Review and an Assessment of Evidentiary Value in Research on the Psychological Consequences of Scarcity,” *Proceedings of the National Academy of Sciences*, 118 (44), e2103313118.
- Open Science Collaboration (2015), “Estimating the Reproducibility of Psychological Science,” *Science*, 349 (6251), aac4716.
- Park, Hanyong, JaeHwan Kwon, and Rajesh Bagchi (2024), “Is ‘4 for \$16’ Better than ‘4 for \$15.30’? The Price Divisibility Effect in Multipack Purchases,” *Journal of Consumer Research*, 51 (3), 455–73.
- Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn (2011), “False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant,” *Psychological Science*, 22 (11), 1359–66.
- Simons, Daniel J., Yuichi Shoda, and D. Stephen Lindsay (2017), “Constraints on Generality (COG): A Proposed Addition to All Empirical Papers,” *Perspectives on Psychological Science*, 12 (6), 1123–8.
- Simonsohn, Uri (2015), “Small Telescopes: Detectability and the Evaluation of Replication Results,” *Psychological Science*, 26 (5), 559–69.
- Sterling, Theodore D. (1959), “Publication Decisions and Their Possible Effects on Inferences Drawn from Tests of Significance—or Vice Versa,” *Journal of the American Statistical Association*, 54 (285), 30–4.
- Urminsky, Oleg and Berkeley J. Dietvorst (2024), “Taking the Full Measure: Integrating Replication into Research Practice to Assess Generalizability,” *Journal of Consumer Research*, 51 (1), 157–68.
- Yarkoni, Tal (2022), “The Generalizability Crisis,” *Behavioral and Brain Sciences*, 45, e1.