NINETY-NINE CENTS: PRICE-ENDING BIAS IN LANGUAGE MODELS IS LEARNED, NOT PERCEIVED

Anonymous authors
Paper under double-blind review

ABSTRACT

Language models respond systematically to price endings: they are much more likely to buy a product at $4.99 than at $5.00, echoing the left-digit bias documented in human consumers. We measure this effect with a within-price placebo design that compares the $4.99–$5.00 change with nine otherwise identical one-cent price increases that do not cross a dollar boundary. The nine placebo steps establish a baseline, isolating the additional effect at the dollar boundary. Across 683,559 cells spanning 21 models, the corrected left-digit effect is -3.25 percentage points $[-5.64, -0.99]$, while a matched design that removes the roundness confound by construction gives -6.14 $[-8.36, -3.86]$. Several experiments then probe whether the bias is perceptual (arising from digit-level processing akin to humans), or semantic (reflecting learned associations with conventional endings). Purchase probability responds to price endings with a graded pattern that favours round endings rather than dropping discretely at the dollar boundary, and the ending profile correlates with corpus frequency. Stating a fictional pricing convention in a single sentence shifts the preferred ending by 36.02 percentage points, roughly ten times the native 99-cent effect. The same design applied to non-price quantities produces domain-specific patterns instead of a uniform digit effect. Rendering prices as images with progressively smaller cent digits produces a perceptual gradient reaching -18.76 points, but this is entirely driven by one of 3 vision models tested. These results suggest that price-ending effects in language models resemble human left-digit bias in their functional form, but reflect learned semantic associations that can be substantially revised by in-context information.

1 INTRODUCTION

Retail prices cluster at nine-endings. In three public pretraining corpora, .99 price endings account for 33.6 percent of all dollar-and-cents prices, and the three most common endings (.99, .00, .95) together account for 68.6 percent (Section 4.2). The standard economic explanation for this pattern is left-digit bias: consumers read prices from left to right, place disproportionate weight on the leftmost digit, and partially discard the cents. A price of $2.99 is encoded as roughly "$2-and-something" rather than "almost $3," so a one-cent increase to $3.00 produces a perceived jump that far exceeds its arithmetic magnitude.

Thomas & Morwitz (2005) formalised this account and established its empirical signature: the demand drop occurs specifically when the leftmost digit changes, not when the price merely increments. When a model assigns a much higher purchase probability to $4.99 than to $5.00, the natural interpretation is that the model has inherited a human perceptual bias. An alternative interpretation is that the model has learned from its training text about which price endings are characteristic of set retail prices. These two accounts make identical predictions at the dollar boundary but differ in their implications for the rest of the price distribution.

To measure left-digit bias, we can compare demand just below and just above a dollar boundary. However, this comparison confounds two effects: a left-digit change and a shift to the ending .00 (which is both the roundest and one of the most common endings). Our design addresses this by embedding the dollar-boundary comparison within a set of nine structurally identical one-cent steps. All ten steps raise the price by one cent and change the final digit from nine to zero. The nine placebo

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

steps establish a baseline response to a 1-cent change, and the excess of the tenth step over that baseline measures the dollar-boundary effect.

Contributions.

1. **A placebo design for isolating the dollar-boundary effect.** We develop a simple within-price placebo design that separates the dollar-boundary effect from the general response to a one-cent price increase and a change from a nine-ending to a zero-ending (Section 3).
2. **Evidence for learned semantic conventions.** Several results suggest that the mechanism is semantic instead of perceptual. A single sentence describing a fictional pricing convention shifts the preferred ending by more than thirty percentage points (Section 4.1). The ending profile is correlated with corpus frequency (Section 4.2), and the effect changes sign across non-price domains in ways consistent with domain-specific conventions (Section 4.3). Models accurately compute one-cent price differences, and the effect persists when prices are written in words (Section 4.4).
3. **A possible perceptual channel in vision.** When prices are rendered as images, and the cent digits are progressively reduced in size, one of 3 vision models shows a monotonic perceptual gradient (Section 5).
4. **Identification of the roundness confound.** We develop a matched design to identify the left-digit effect without explicit roundness correction, and we show that accounting for the roundness confound is essential: the standard dollar-boundary comparison can misrepresent both the magnitude and direction of the underlying effect (Section 6).
5. **Prompt sensitivity exceeds model variation.** The measured price-ending effect varies more across twelve prompt phrasings than across fifteen models (Section 7).

2 Related work

The left-digit effect has been studied extensively in both laboratory and field settings. Schindler & Kirby (1997) documented the prevalence of rightmost-digit patterns in advertised prices and argued that they exploit consumers' tendency to truncate instead of rounding. Stiving & Winer (1997) used scanner data to show that demand is more sensitive to left-digit changes than to equivalent price changes that leave the leftmost digit unchanged. Thomas & Morwitz (2005) provided the clearest experimental evidence: across five studies, the demand response appears only when the leftmost digit changes. Anderson & Simester (2003) found supporting evidence in field experiments with catalogue prices. Lacetera et al. (2012) found sharp discontinuities at 10,000-mile odometer thresholds in used-car auctions. List et al. (2023) estimated the left-digit share of the per-dollar demand decline in ride-hailing, and Strulov-Shlain (2023) developed a structural model of left-digit pricing using retail scanner data. Sokolova et al. (2020) showed that the bias is stronger in stimulus-based evaluations, and Pope & Simonsohn (2011) documented round-number effects in non-price settings, motivating our treatment of roundness as a distinct phenomenon.

Language models have been proposed as simulated experimental subjects for economic research. Horton et al. (2023) introduced the concept and demonstrated that models can replicate classical experimental findings. Gui & Toubia (2023) raised causal concerns about this approach, arguing that correlations in model outputs need not reflect the same mechanisms as in human data. Hagendorff et al. (2023) showed that human-like reasoning biases present in one model generation can disappear in the next. Fish et al. (2024) studied language models as pricing agents and found evidence of algorithmic collusion. Our results identify a complementary failure mode: a model can reproduce a human demand anomaly in its functional form while the underlying mechanism is entirely different.

3 Design and measurement

3.1 Task

For every API call, we give the model one product (Appendix A.1), one persona (Appendix A.2), and one price, and request a simple binary decision on whether to buy or to walk away. The model then replies with just one word, either YES or NO. The complete prompt is given in Appendix A.3. We

calculate the purchase probability in two ways. For models that provide log-probabilities, we obtain $P(\text{buy})$ directly from the first-token probabilities:

$$P(\text{buy}) \;=\; \frac{\sum_{t\in\text{YES}} e^{\ell_t}}{\sum_{t\in\text{YES}} e^{\ell_t} + \sum_{t\in\text{NO}} e^{\ell_t}}, \tag{1}$$

where ℓ_t is the log-probability of token t and the YES and NO sets contain the relevant surface forms. For models without log-probabilities, we estimate purchase probability from 12 repeated draws per cell. Of the 21 models used in this paper, 10 expose log-probabilities (Appendix A.4).

3.2 Price grid

Different prices at which the model is asked to make a buy or not decision are organised into blocks. Each block fixes the model, product, persona, and dollar anchor b, and varies only the price between $\$b.00$ and $\$(b+1).00$. There are 31 prices within each block: for each decade $d = 0, \ldots, 9$, we include the endings $.d0$, $.d5$, and $.d9$, followed by $\$(b+1).00$. This creates ten one-cent transitions from a nine-ending to the next zero-ending:

$$\underbrace{.09 \to .10 \quad .19 \to .20 \quad \cdots \quad .89 \to .90}_{\text{nine placebo steps}} \qquad \underbrace{.99 \to +1.00}_{\text{treatment step}}$$

All ten transitions increase the price by one cent and change the final digit from nine to zero. The only difference is that the final transition also changes the dollar digit. The nine placebo steps therefore provide a baseline for the response to a one-cent increase, and the excess at the tenth step isolates the dollar-boundary effect.

3.3 Estimator

Let y_{ib} denote the purchase probability at price point i in block b, and let $x_{ib} \in [0, 1]$ denote the position within the dollar. We remove block differences by subtracting each block's mean, then estimate and remove the common within-dollar price trend:

$$\hat{\beta} \;=\; \frac{\sum_b \sum_i \tilde{x}_{ib}\, \tilde{y}_{ib}}{\sum_b \sum_i \tilde{x}_{ib}^2}, \tag{2}$$

where tildes denote block-demeaned values. The residual $\hat{u}_{ib} = \tilde{y}_{ib} - \hat{\beta}\tilde{x}_{ib}$ captures variation not explained by the block or trend. For each ending e, we average the residual across multiple blocks:

$$\hat{\mu}_e \;=\; 100 \times \text{mean}_{\{(i,b)\,:\,\text{offset}_{ib}=e\}}\, \hat{u}_{ib}, \tag{3}$$

giving the effect of each ending in percentage points. The jump at step s is $\delta_s = \hat{\mu}_{(10s) \bmod 100} - \hat{\mu}_{10s-1}$. Figure 1 illustrates the procedure.

Inference. We exclude blocks whose within-block range in purchase probability falls below one percentage point, as these are uninformative about ending effects. All intervals are block bootstraps over 400 resamples, reporting the 2.5th and 97.5th percentiles; * marks intervals excluding zero. Applying the Benjamini–Hochberg correction across all 12 arm-level tests, 7 remain significant at the 5% level (Appendix D).

3.4 Experiments

The core sweep (E1) runs the full price grid on 15 models from ten families. Five experiments are conducted to probe the mechanism: stated convention (E10), corpus frequency (E5), non-price domains (E3), number formatting (E6), and explicit reasoning (E8). A vision experiment (E7) tests the perceptual channel. Two identification experiments address the roundness confound: E0 uses a matched design with two-digit prices; E0b extends this to three-digit prices. A prompt-phrasing experiment (E2) measures sensitivity to wording. Table 1 summarises the data collected across all thirteen experiments, and Appendix A provides full details.

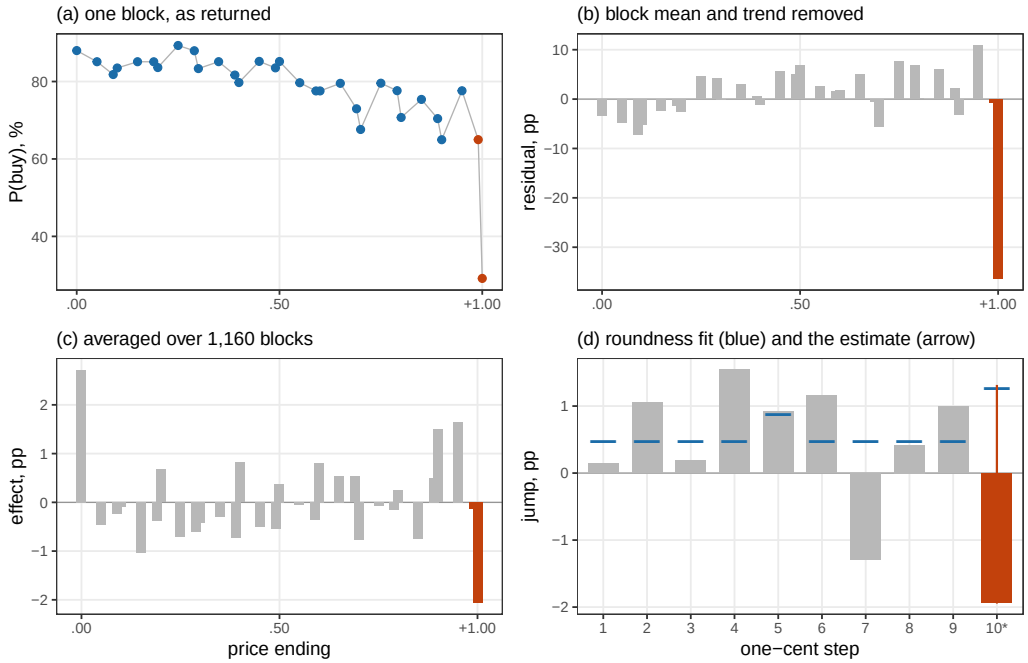

Figure 1: Construction of the estimate. (a) One block: 31 prices from $3.00 to $4.00. (b) After removing the block mean and within-dollar price trend. (c) Averaged over 1,160 blocks: the ending profile. (d) The ten one-cent steps.

Table 1: Data collected.

Experiment	Code	Models	Cells	Calls
Dollar boundary type	E0	9	93,555	93,555
Digit position	E0b	9	64,152	64,152
Core sweep	E1	15	55,800	260,400
Stated convention	E10	5	15,840	15,840
Scaled prices	E1b	5	7,440	7,440
Newest models	E1f	5	3,720	44,640
Prompt phrasing	E2	5	44,640	44,640
Non-price magnitudes	E3	5	223,200	223,200
Currency and locale	E4	5	18,600	18,600
Number formatting	E6	4	14,880	14,880
Rendered price tags	E7	4	138,942	138,942
Explicit reasoning	E8	5	930	2,976
Arithmetic probes	E9	15	1,860	8,680
Total			683,559	937,945

A cell is one condition at one price, averaging the repeated draws that a sampled read-out needs. Calls is the number of API requests behind those cells.

3.5 THE ENDING PROFILE

The ending profile shows how much each of the thirty price endings raises or lowers purchase probability, after removing the block level and within-block trend. Figure 1(c) plots the profile from the core sweep. The profile has two key features. First, models respond to all price endings in a graded pattern that favours round endings. The three endings with the largest effects are .00 (2.72 pp), .95 (1.65 pp), and .90 (1.49 pp), all of which are round, and the nine placebo steps average 0.57 pp, meaning that moving from a nine-ending to the next round ending slightly increases demand even within a dollar. A pure threshold account, in which only the change in the leftmost digit matters, would instead predict a flat profile with a single downward step at the dollar boundary. Second, step

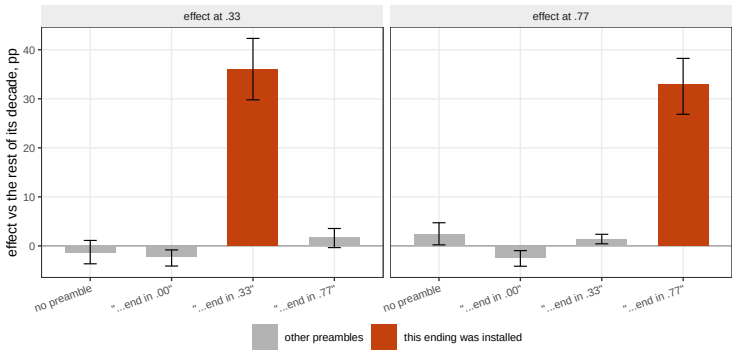

Figure 2: Effect of stating a pricing convention. Each panel shows the effect of one ending, with each bar corresponding to a different preamble. The highlighted bar indicates the ending named in the preamble.

10 produces a -1.93 pp change, despite ending at a whole dollar, the roundest price on the grid. Fitting the roundness pattern to the nine placebo steps predicts a $+1.32$ pp change at step 10. The difference between this prediction and the observed change provides our estimate of the additional left-digit effect, which we describe in detail in Section 6. Per-model estimates are reported in Table 7 (Appendix B.1).

4 Mechanism: A Learned Convention

We describe five experiments that test whether the ending profile reflects a perceptual distortion or a learned association between endings and retail pricing.

4.1 Stated Convention

Here, we add a single sentence to the prompt stating that retail prices in the country conventionally end in a particular value, such as .77 or .33. We also add .33 and .77 to the price grid and measure the effect of each ending relative to other endings within the same decade.

Without the sentence, .33 has an effect of -1.46 and .77 has an effect of 2.32 (neither ending standing out). Stating the .33 convention raises the effect of .33 to 36.02 $[29.79,\ 42.31]$, while .77 remains at 1.27. Conversely, stating the .77 convention raises the effect of .77 to 33.02 $[26.84,\ 38.23]$, while .33 remains at 1.65 (Figure 2; Tables 26, 27, and 28 in Appendix B.11).

Thus, a single sentence produces a shift of roughly thirty to thirty-six percentage points, about ten times the native 99-cent effect. The shift is concentrated on the ending named in the sentence rather than appearing across endings, ruling out a general increase in demand. If the 99-cent effect were a hard-wired perceptual distortion, it should be robust to a one-sentence prompt change. Instead, the ease with which context reshapes the ending preference suggests that the native effect is itself a learned association of the same kind, merely reflecting the dominant convention in the training data.

4.2 Corpus Frequency

We count literal price strings in three public pretraining corpora, C4, Dolma, and RedPajama, using the infini-gram index (Liu et al., 2025). The infini-gram index provides efficient access to n-gram frequencies in large text corpora, allowing us to measure how often each price ending appears in training text. Price endings are observed to be highly concentrated: .99 accounts for 33.6 percent of all posted prices, followed by .00 at 19.2 percent and .95 at 15.9 percent. These endings alone account for 68.6 percent of all posted prices. The ordering is also consistent across corpora, with pairwise Spearman correlations ranging from 0.968 to 0.989.

Figure 3(b) shows that the ending profile from the core sweep is positively correlated with posting frequency, with a Spearman correlation of 0.333 across 30 endings. This means that endings with larger effects on the probability of buying tend to be more common in web text, consistent with preferences reflecting learned associations from the training distribution (Table 29 in Appendix B.12).

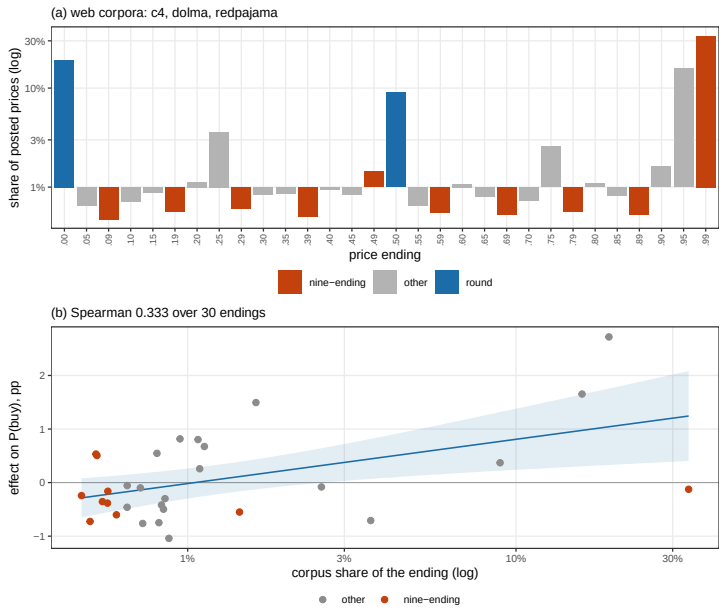

Figure 3: Price-ending frequency and model preferences. (a) Frequency of each ending in web text. (b) Model ending effect plotted against posting frequency. Spearman $\rho = 0.333$.

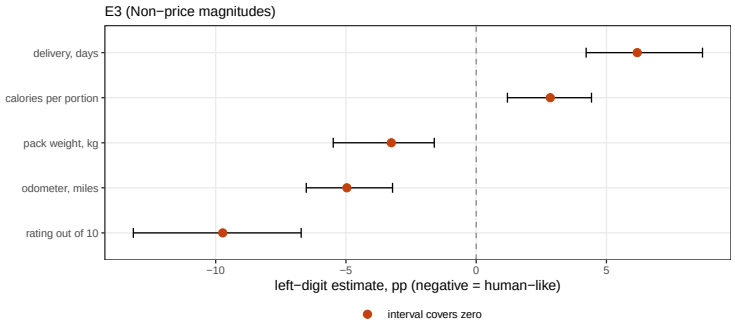

Figure 4: Effects of dollar-boundary-like transitions across non-price domains.

4.3 Non-price domains

If the effect reflected a response to the digits themselves, the same numerical transition should produce a similar response in other quantities. We apply the same decade structure to 5 non-price quantities: odometer readings, calorie counts, package weights, delivery times, and ratings. All 5 domains show significant effects, but the effects differ in sign across domains (Figure 4). The same numerical change that produces a negative response in prices produces a positive response in two other domains. This pattern is difficult to reconcile with a universal digit-level mechanism, but consistent with domain-specific conventions learned from training text. Prices ending in .99 signal deliberate retail pricing, while round odometer readings or calorie counts carry their own semantic associations. The model's response to a numerical boundary therefore depends on the domain it has been prompted with, not on the digits alone (Tables 14 and 15 in Appendix B.5).

4.4 Arithmetic and tokenisation

We also use a simple exercise to test whether the effect could reflect basic arithmetic or tokenisation. When asked which of two adjacent prices is lower, models answer correctly in 99.8% of trials (Table 25 in Appendix B.10). They also correctly compute the one-cent difference between $4.99



Figure 5: The images shown to the vision models, at the two prices on either side of a dollar boundary. Tags are rendered at 512×512 pixels and cached.

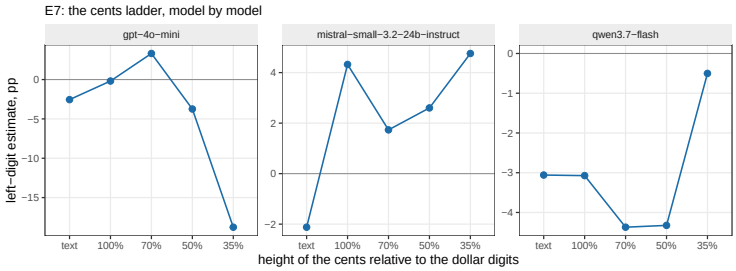

Figure 6: Effect of reducing the visual size of the cents. Each panel shows the left-digit estimate at four relative cent sizes. Only `gpt-4o-mini` shows a monotonic gradient.

and \$5.00. The ending profile therefore cannot be explained by a failure to represent the price difference. Presenting the same prices in 5 formats (standard, spaced, digit-separated, ISO, spelled out in words) gives a pooled estimate of -0.26 $[-3.42, 2.99]$. The word rendering, which removes the digit structure entirely, gives 6.28. The profile therefore persists across representations and cannot be attributed to tokenisation (Tables 18 and 19 in Appendix B.7).

5 A PERCEPTUAL CHANNEL IN VISION

So far, the evidence points to semantic mechanisms that are learned from text. Images provide a complementary test: unlike text prompts, visual displays allow us to directly manipulate the perceptual salience of different parts of a price. We therefore ask whether making the cents less visually prominent increases the dollar-boundary effect. Thus, we visually render each price as a shelf tag (Figure 5) and vary the cents height to full size, 70%, 50%, and 35%. If the effect is partly perceptual, making the cents less prominent should make the dollar-boundary effect larger.

Pooled across 3 vision models, the gradient is 5.65 percentage points per unit reduction in relative cents height $[-2.05, 14.95]$. However, the gradient is driven almost entirely by one model: in `gpt-4o-mini`, the estimate shifts from -0.19 at 100% to -18.76 at 35%, a gradient of 25.53 $[0.88, 50.45]$. The corresponding estimates are flat in the other models (Figure 6; Tables 20, 22, and 21 in Appendix B.8). These results provide some evidence for a perceptual channel, but it is not a general property of vision models.

6 THE ROUNDNESS CONFOUND

As discussed before, step 10 (\$$b$.99 to \$$(b+1)$.00) differs from the other steps in two aspects: it crosses the dollar boundary and ends at .00, the roundest price on the grid. Because models prefer round endings (Section 3.5), the raw step-10 comparison combines two opposing effects. Crossing

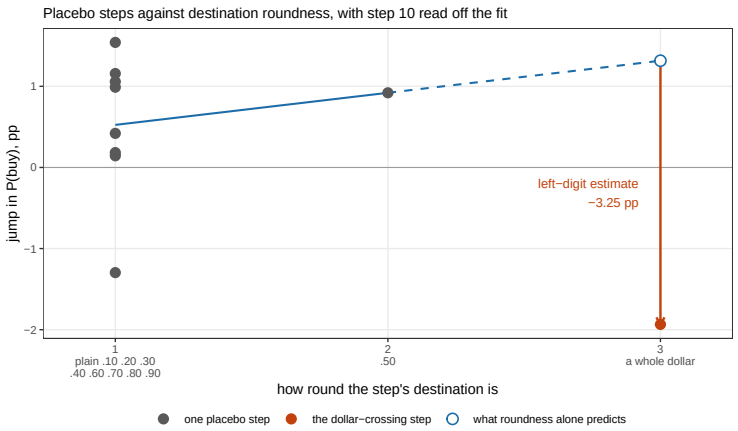

Figure 7: The roundness correction on the core sweep. Grey points are placebo steps; the line is fitted to the nine placebos and extrapolated to the whole-dollar destination. The arrow shows the corrected left-digit estimate.

the dollar boundary lowers purchase probability through the left-digit effect, while arriving at .00 raises it through the preference for round endings. In the core sweep, the raw step-10 jump is -2.50 pp. The attraction to the round \$X.00 ending offsets part of the left-digit effect, so the raw comparison attenuates the magnitude of the underlying left-digit effect. In other words, the observed drop at the dollar boundary is smaller than the left-digit effect would be in the absence of the roundness preference (Table 38 in Appendix C).

Roundness correction. We first consider a simple method to remove the roundness confound by assigning each destination a roundness score ρ_e: 3 for a whole dollar, 2 for .50, 1 for a multiple of ten, and 0 otherwise. We fit $\delta_s = a + b\,\rho_{d(s)} + \varepsilon_s$ to the nine placebo steps and take the residual at step 10 as the corrected estimate. This gives -3.25 pp $[-5.64,\ -0.99]$ across 1,160 blocks. Mapping this to the inattention parameter of Strulov-Shlain (2023) gives $\hat{\theta} = 0.76$ $[0.24,\ 1.32]$, above the range of 0.15 to 0.25 estimated from retail scanner data (Strulov-Shlain, 2023), the 0.30 from used-car auctions (Lacetera et al., 2012), and the ≈ 0.5 from ride-hailing (List et al., 2023). The comparison is indirect because the human parameter is identified from equilibrium prices rather than stated intentions. The estimate is spread across models rather than driven by a small subset (Table 8 in Appendix B.1). Two alternative corrections (corpus frequency and a nonparametric monotone bound) yield similar estimates (Appendix C).

Figure 7 illustrates this correction. The nine placebo steps establish the relationship between destination roundness and the size of the step response. We extrapolate this relationship to the whole-dollar destination and interpret the remaining difference between the observed and predicted step-10 responses as the left-digit component. We find no systematic pattern by scale, model family, or reasoning capability (Table 41; Table 40 in Appendix C).

Matched design. An alternative approach is to remove the roundness confound by construction, rather than estimating and subtracting it. With two-digit dollar prices, the dollar figure can increase without changing the leftmost digit (\$12.99 → \$13.00, the `same` condition) or together with the leftmost digit (\$19.99 → \$20.00, the `roll` condition). Both end at .00, so their difference isolates the left-digit effect:

$$\widehat{\mathrm{LD}} \;=\; \delta_{10}\big|_{\texttt{roll}} \;-\; \delta_{10}\big|_{\texttt{same}} \tag{4}$$

In the `same` condition, the dollar figure increases but the leftmost digit does not, and purchase probability rises by 3.27 pp. In the `roll` condition, the leftmost digit also changes, and purchase probability falls by -2.87 pp. The difference is -6.14 $[-8.36,\ -3.86]$, with a negative estimate in 7 of 9 models (Tables 30, 31, and 32 in Appendix B.13). The effect therefore depends on whether the leftmost digit changes, consistent with the criterion established in human experiments by Thomas & Morwitz (2005). This comparison is not available with single-digit dollar prices, where \$4.99 → \$5.00 necessarily changes both the dollar figure and the leftmost digit. In Appendix B.14, we also discuss a three-digit extension that uses prices around \$200 to further refine the estimates.

7 EXTENSIONS

Prompt phrasing. Repeating the design under 12 alternative phrasings gives a pooled estimate of -8.91 $[-11.04, -6.93]$. Of the 12 templates, 10 produce individually significant negative estimates, and none produce a positive estimate. Estimates range from 0.13 to -15.94 across templates, which is larger than the corresponding range across the 15 models in the core sweep. Thus, the estimated effect depends substantially on how the question is phrased, and reporting a single value per model without specifying the prompt can obscure this source of variation (Tables 12 and 13; Figure 8 in Appendix B.4).

Other settings. The effect is also present across currencies and alternative experimental settings. Across 5 currencies, the pooled estimate is -2.73 $[-5.04, -0.15]$; Japanese yen, which is quoted in whole numbers without a cents field, gives -3.42 $[-7.99, 1.71]$ (Tables 16 and 17; Figure 9 in Appendix B.6). Requiring a chain of thought before the answer gives -7.68 $[-56.93, 38.27]$ across 20 blocks (Tables 23 and 24 in Appendix B.9). The newest closed models, sampled 12 times per cell, give -15.54 $[-38.70, 9.16]$ across 48 blocks (Table 9 in Appendix B.2).

8 DISCUSSION

Summary of findings. Models exhibit a smooth ending profile, with graded preferences for round and common endings (Section 3.5). Multiple pieces of evidence point to learned conventions: a single sentence of context can induce a thirty-point preference for an arbitrary ending (Section 4.1), the profile correlates with corpus frequency (Section 4.2), and the same design applied to non-price quantities yields domain-specific patterns rather than a universal digit effect (Section 4.3). Models also accurately compute the underlying one-cent difference (Section 4.4). Although a perceptual channel is present in vision, its gradient is concentrated in only one of 3 models (Section 5). In a matched design, we identify a left-digit effect of -6.14 $[-8.36, -3.86]$ after accounting for roundness. Finally, prompt phrasing changes the measured effect more than model identity (Section 7).

Implications for simulated consumers. Reproducing a human demand anomaly does not imply reproducing the process that generates it. Models reproduce the defining two-digit price-ending pattern, but the underlying mechanism differs: simulated demand responds to price endings, yet responds about 10x more to a sentence describing conventions than to the native 99-cent effect.

Implications for bias evaluation. The standard .99 versus .00 comparison confounds the left-digit change with the roundness of the destination. Our placebo and matched designs separate them. The sensitivity of the measured effect to prompt phrasing, which exceeds the variation across models, means that a benchmark score obtained under one wording may not generalise to another. Evaluations of pricing bias in language models should therefore report the prompt alongside the estimate.

Limitations. All models in our sample are instruction-tuned; base models may respond differently to price endings. The design measures stated purchase intentions from a single binary choice and not the actual purchasing behaviour. The corpus-frequency analysis uses publicly indexed corpora that may not match each model's actual training mixture.

9 CONCLUSION

Language models exhibit a 99-cent effect that resembles the defining feature of left-digit bias: demand drops when the leftmost digit changes, but not when the dollar figure increases by one cent within the same dollar. A matched design estimates this effect at -6.14 percentage points after removing the roundness confound by construction. However, the evidence does not support a perceptual mechanism. The ending profile is graded, correlates with corpus frequency, reverses sign in non-price domains, and shifts by roughly 30 percentage points in response to a single sentence naming a fictional pricing convention. Models also compute the underlying one-cent difference accurately, and a vision perceptual gradient appears in only one of 3 models tested. These results suggest that the 99-cent effect reflects a learned pricing convention: models reproduce the behavioural pattern, but the convention underlying their response can be reshaped by context.

ETHICS STATEMENT

This work queries commercial language model APIs with synthetic retail vignettes. No human subjects were involved, and no personal data was collected. The products, prices, and personas are all invented.

AI USE STATEMENT

In this work, we used generative AI tools (Claude, GPT) for writing and debugging parts of code and for offering suggestions on draft text. We have not used generative AI tools for formulating research questions, designing experiments, or drawing scientific conclusions, and tasks such as data collection from human subjects and mathematical proofs do not apply to this work. Additionally, we used generative AI tools for checking writing style and for identifying any inconsistencies across the manuscript. We have reviewed all AI-assisted work: all code was tested against known outputs before use, and all suggested text was rewritten or verified by the author. We take responsibility for the final content of this work, including all text, claims, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The complete experimental specification, including all prompt templates, price grids, product lists, persona descriptions, and model identifiers, is given in Appendix A. The processed data panels, analysis code, and all output files are available in the supplementary material and at `https://anonymous.4open.science/r/LLM_Left_Digit_Bias_Codes-4680`. Block bootstrap confidence intervals use 400 replications with a fixed seed for exact reproducibility.

REFERENCES

Eric Anderson and Duncan Simester. Effects of $9 price endings on retail sales: Evidence from field experiments. *Quantitative Marketing and Economics (QME)*, 1(1):93–110, March 2003. doi: 10.1023/A:1023581927405. URL https://ideas.repec.org/a/kap/qmktec/v1y2003i1p93-110.html.

Sara Fish, Yannai A. Gonczarowski, and Ran I. Shorrer. Algorithmic collusion by large language models. Papers 2404.00806, arXiv.org, Mar 2024. URL https://ideas.repec.org/p/arx/papers/2404.00806.html.

George Gui and Olivier Toubia. The challenge of using llms to simulate human behavior: A causal inference perspective. 2023. doi: 10.2139/ssrn.4650172. URL http://dx.doi.org/10.2139/ssrn.4650172.

Thilo Hagendorff, Sarah Fabi, and Michal Kosinski. Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in chatgpt. *Nature Computational Science*, 3:1–6, 10 2023. doi: 10.1038/s43588-023-00527-x.

John J Horton, Apostolos Filippas, and Benjamin S Manning. Large language models as simulated economic agents: What can we learn from homo silicus? Working Paper 31122, National Bureau of Economic Research, April 2023. URL http://www.nber.org/papers/w31122.

Nicola Lacetera, Devin G. Pope, and Justin R. Sydnor. Heuristic thinking and limited attention in the car market. *American Economic Review*, 102(5):2206–36, May 2012. doi: 10.1257/aer.102.5.2206. URL https://www.aeaweb.org/articles?id=10.1257/aer.102.5.2206.

Amit Arnold Levy and Mor Geva. Language models encode numbers using digit representations in base 10. In Luis Chiruzzo, Alan Ritter, and Lu Wang (eds.), *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pp. 385–395, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics. ISBN 979-8-89176-190-2. doi: 10.18653/v1/2025.naacl-short.33. URL https://aclanthology.org/2025.naacl-short.33/.

John A List, Ian Muir, Devin Pope, and Gregory Sun. Left-digit bias at lyft. *The Review of Economic Studies*, 90(6):3186–3237, 11 2023. ISSN 0034-6527. doi: 10.1093/restud/rdad014. URL https://doi.org/10.1093/restud/rdad014.

Jiacheng Liu, Sewon Min, Luke Zettlemoyer, Yejin Choi, and Hannaneh Hajishirzi. Infini-gram: Scaling unbounded n-gram language models to a trillion tokens, 2025. URL https://arxiv.org/abs/2401.17377.

Devin Pope and Uri Simonsohn. Round numbers as goals: Evidence from baseball, sat takers, and the lab. *Psychological Science*, 22(1):71–79, 2011. ISSN 09567976, 14679280. URL http://www.jstor.org/stable/40984610.

Robert M. Schindler and Patrick N. Kirby. Patterns of rightmost digits used in advertised prices: Implications for nine-ending effects. *Journal of Consumer Research*, 24(2):192–201, 1997. ISSN 00935301, 15375277. URL http://www.jstor.org/stable/10.1086/209504.

Tatiana Sokolova, Satheesh Seenivasan, and Manoj Thomas. The left-digit bias: When and why are consumers penny wise and pound foolish? *Journal of Marketing Research*, 57(4): 771–788, 2020. doi: 10.1177/0022243720932532. URL https://doi.org/10.1177/0022243720932532.

Mark Stiving and Russell S. Winer. An empirical analysis of price endings with scanner data. *Journal of Consumer Research*, 24(1):57–67, 1997. ISSN 00935301, 15375277. URL http://www.jstor.org/stable/10.1086/209493.

Avner Strulov-Shlain. More than a penny's worth: Left-digit bias and firm pricing. *The Review of Economic Studies*, 90(5):2612–2645, 10 2023. ISSN 0034-6527. doi: 10.1093/restud/rdac082. URL https://doi.org/10.1093/restud/rdac082.

Manoj Thomas and Vicki Morwitz. Penny wise and pound foolish: The left-digit effect in price cognition. *Journal of Consumer Research*, 32(1):54–64, 2005. ISSN 00935301, 15375277. URL http://www.jstor.org/stable/10.1086/429600.

Appendix

A Materials

A.1 Products

Eight everyday goods spanning \$5 to \$27. The MINI grid uses the first four. Each product has a situation making the outside option costly, no brand or quality signal, and a flat descriptor.

ID	Item	Descriptor	Situation	Ref
latte	a 12 oz latte	made to order, oat milk available	stopped at a cafe, want a coffee	\$5
sandwich	a chicken salad sandwich	from the chiller	1 pm, skipped breakfast	\$7
notebook	a hardcover A5 notebook	200 lined pages	stationery shop, notebook full	\$8
cable	a USB-C charging cable	fast charging, 1 yr warranty	electronics shop, old cable broken	\$12
pizza	a large pepperoni pizza	ready in 15 min	8 pm, have not cooked	\$16
tshirt	a plain cotton t-shirt	mid-weight, your size	clothing shop	\$18
umbrella	a compact windproof umbrella	fits in a bag	raining, no umbrella	\$22
headphones	wired in-ear headphones	in-line mic, decent reviews	headphones broke this morning	\$27

E0 uses 17 products with two-digit reference prices ($21–$92), each posted at a `roll` anchor (leftmost digit changes) and a `same` anchor (it does not). E0b uses 8 products with three-digit reference prices ($210–$515), each posted at three anchors (`flat`, `inner`, `lead`).

A.2 PERSONAS

Eleven personas are used in total. Five are used in the core grid (P1–P5); four more (P7–P10) are added for the enlarged grids; two special conditions (P0, P6) appear in E2 only.

ID	Label	Delib.	Text given to the model
Core personas (CORE and MINI grids)			
P1	student, careful	high	21 yr old student, tight budget, keeps a spending log
P2	student, impulsive	low	20 yr old student, little spare money, decides quickly
P3	professional, deal-conscious	high	38 yr old professional, comfortable salary, dislikes overpaying
P4	parent, distracted	low	31 yr old parent of toddler, tired, in a hurry
P5	retiree, careful	high	68 yr old retiree, fixed pension, thinks carefully
Extended personas (E3, E7, E0, E0b)			
P7	freelancer, irregular income	high	29 yr old freelancer, mental budget per category
P8	senior manager, price-insensitive	low	45 yr old senior manager, rarely attends to small prices
P9	schoolteacher, weighs value	high	52 yr old schoolteacher, weighs up whether worth the money
P10	gig worker, impulsive	low	26 yr old gig worker, buys on impulse
Special conditions (E2 only)			
P0	no persona	—	(prompt substitutes "You are an ordinary shopper.")
P6	student, time pressure	load	P2 + "You have ninety seconds before your bus leaves."

A.3 PROMPT TEMPLATES

System prompt (identical across all choice arms):

> You are role-playing one specific consumer making a real purchase decision in a shop. Decide the way that consumer would, given their circumstances and the price. Reply with exactly one word: YES if you buy it, NO if you do not. No explanation, no punctuation.

Canonical template (T01), used by every arm except E2:

> About you: {persona}
> Situation: {setting}.
> The item: {name}, {descriptor}. Price: {price}
> You have no coupon and you have not checked prices anywhere else today. If you do not buy it now, you go without it.
> Do you buy it at this price?

All twelve templates. E2 sweeps templates T01–T12.

ID	What it varies
T01	Canonical. Structured fields, second person, outside option stated.
T02	Flowing prose rather than labelled fields.

ID	What it varies
T03	Third person ("Does this shopper buy it?").
T04	Price stated first, before the item and persona.
T05	Single flowing paragraph, constraints inline.
T06	Third person, "Consider a shopper".
T07	Terse labels ("You / Where / What / How much / Buy?").
T08	Outside option removed.
T09	"Imagine you are the following person" framing.
T10	Bracketed machine-readable fields.
T11	Prose with the outside option restated at the end.
T12	Compact prose with an em-dash price line.

A.4 MODELS

Twenty-one models across ten families are used.

Model	Tier	Arms
Precision (10): log-probabilities, 1 call/cell		
`openai/gpt-4o`	high	E1, E7, E9
`openai/gpt-4o-mini`	low	all factorial
`openai/gpt-oss-20b`	low	all factorial, E0, E0b
`qwen/qwen3-235b-a22b-2507`	high	all factorial, E0, E0b
`qwen/qwen3-30b-a3b-2507`	mid	all factorial, E0, E0b
`qwen/qwen3.5-9b`	low	E1, E9, E0, E0b
`deepseek/deepseek-v4-flash`	mid	all factorial, E0, E0b
`deepseek/deepseek-chat-v3-0324`	mid	E1, E9, E0, E0b
`qwen/qwen3.7-flash`	mid	E1, E7, E9, E0, E0b
`mistralai/mistral-small-3.2-24b`	mid	all factorial, E7, E0, E0b
Sampled (6): no log-probabilities, 12 calls/cell		
`openai/gpt-oss-120b`	high	E1, E9
`qwen/qwen3-235b-thinking-2507`	high	E8
`qwen/qwen3-32b`	mid	E1, E9
`qwen/qwen3-14b`	mid	E1, E9
`google/gemma-3-27b-it`	mid	E1, E9
`google/gemma-4-31b-it`	mid	E1, E9
Frontier (5): newest closed models, 12 calls/cell		
`openai/gpt-5.6-sol`	high	E1f
`openai/gpt-5.6-luna`	low	E1f
`anthropic/claude-sonnet-5`	high	E1f
`google/gemini-3.7-flash`	high	E1f
`google/gemini-3.5-flash-lite`	low	E1f

A.5 NAMED GRID SETS

Grid	Products	Anchors	Prices/block	Personas	Cells/model
CORE	8	3	31	5 (P1–P5)	3,720
MINI	4	2	31	3 (P1, P2, P5)	744

B EXPERIMENT RESULTS

Each experiment keeps the design described in Section 3 fixed. The subsections below describe the per-condition and per-model estimates for each experiment.

Table 7: E1 (Core sweep): estimates by model.

Model	Blocks	Slope	Left-digit	95% CI
gpt-oss-20b	91	-10.22	-18.41	[-39.25, 7.06]
deepseek-v4-flash	83	-6.31	-7.13*	[-12.47, -2.49]
gemma-3-27b-it	27	-15.72	-6.40	[-27.78, 18.01]
gpt-4o-mini	69	-6.40	-6.19	[-14.66, 1.31]
qwen3.7-flash	118	-2.26	-5.03*	[-9.04, -1.58]
deepseek-chat-v3-0324	81	-1.51	-3.97	[-9.41, 0.89]
qwen3-235b-a22b-2507	83	0.76	-3.37	[-9.14, 3.90]
mistral-small-3.2-24b-instruct	111	-4.12	-1.23	[-2.53, 0.27]
qwen3-14b	83	-2.78	-1.12	[-10.04, 7.48]
qwen3-32b	98	-1.19	-0.53	[-7.35, 4.81]
gpt-4o	65	-4.88	-0.51	[-6.27, 6.81]
qwen3.5-9b	117	-3.32	0.38	[-0.85, 1.92]
gemma-4-31b-it	20	-11.33	0.83	[-14.15, 13.94]
gpt-oss-120b	56	-7.37	2.66	[-6.39, 11.73]
qwen3-30b-a3b-instruct-2507	58	-2.97	6.86	[-1.12, 15.32]

Experiment	**Code**	**Models**	**What it varies**
Core sweep	E1	15	Nothing (the baseline).
Newest models	E1f	5	The core design on newest closed models (sampled).
Scaled prices	E1b	5	Prices at $\times 10$ and $\times 100$.
Prompt phrasing	E2	5	12 wordings of the question.
Non-price magnitudes	E3	5	5 non-price domains.
Currency and locale	E4	5	5 currencies including JPY.
Number formatting	E6	4	5 renderings of the price.
Rendered price tags	E7	4	The price as an image at four cents sizes.
Explicit reasoning	E8	5	Chain-of-thought before answering.
Arithmetic probes	E9	15	Which price is lower, and by how much.
Stated convention	E10	5	4 preambles naming a convention.
Dollar boundary type	E0	9	Whether the leftmost digit changes.
Digit position	E0b	9	Which digit boundary a three-digit price crosses.
Corpus counts	E5	n/a	Ending frequency in three public corpora.

B.1 Core sweep (E1)

Table 7 reports the left-digit estimate for each model in the core sweep. Table 8 provides the corresponding ending profiles.

B.2 Newest models (E1f)

Table 9 reports the left-digit estimate for each frontier model.

B.3 Scaled prices (E1b)

Tables 10 and 11 report the estimates by condition and by model.

B.4 Prompt phrasing (E2)

Tables 12 and 13 report the estimates by template and by model. Figure 8 plots the estimate under each template.

Table 8: Left-digit effect by model, core sweep (E1).

Model	Tier	Readout	Blocks	Slope	Left-digit	95% CI
qwen3-235b-a22b-2507	high	logprob	83	0.76	-3.37	[-9.14, 3.90]
gpt-4o	high	logprob	65	-4.88	-0.51	[-6.27, 6.81]
gpt-oss-120b	high	sample	56	-7.37	2.66	[-6.39, 11.73]
gpt-oss-20b	low	logprob	91	-10.22	-18.41	[-39.25, 7.06]
gpt-4o-mini	low	logprob	69	-6.40	-6.19	[-14.66, 1.31]
qwen3.5-9b	low	logprob	117	-3.32	0.38	[-0.85, 1.92]
deepseek-v4-flash	mid	logprob	83	-6.31	-7.13*	[-12.47, -2.49]
gemma-3-27b-it	mid	sample	27	-15.72	-6.40	[-27.78, 18.01]
qwen3.7-flash	mid	logprob	118	-2.26	-5.03*	[-9.04, -1.58]
deepseek-chat-v3-0324	mid	logprob	81	-1.51	-3.97	[-9.41, 0.89]
mistral-small-3.2-24b-instruct	mid	logprob	111	-4.12	-1.23	[-2.53, 0.27]
qwen3-14b	mid	sample	83	-2.78	-1.12	[-10.04, 7.48]
qwen3-32b	mid	sample	98	-1.19	-0.53	[-7.35, 4.81]
gemma-4-31b-it	mid	sample	20	-11.33	0.83	[-14.15, 13.94]
qwen3-30b-a3b-instruct-2507	mid	logprob	58	-2.97	6.86	[-1.12, 15.32]

Readout is first-token log-probabilities where the provider exposes them and the mean of 12 sampled binary answers otherwise.

Table 9: E1f (Newest models): estimates by model.

Model	Blocks	Slope	Left-digit	95% CI
gpt-5.6-sol	9	-27.89	-26.50*	[-40.64, -14.42]
claude-sonnet-5	6	-17.91	-21.70	[-55.33, 10.52]
gemini-3.7-flash	8	-28.92	-16.28	[-60.34, 27.38]
gemini-3.5-flash-lite	12	-22.52	-15.62	[-103.65, 59.06]
gpt-5.6-luna	13	-16.72	-4.57	[-20.06, 6.63]

B.5 Non-price domains (E3)

Tables 14 and 15 report the estimates by domain and by model.

B.6 Currency and locale (E4)

Tables 16 and 17 report the estimates by currency and by model. Figure 9 plots the estimates in each currency.

B.7 Number formatting (E6)

Tables 18 and 19 report the estimates by rendering format and by model.

B.8 Rendered price tags (E7)

Table 20 reports the left-digit estimate at each cents size. Table 21 breaks this out by model, and Table 22 reports the linear trend.

B.9 Explicit reasoning (E8)

Tables 23 and 24 report the estimates by condition and by model.

B.10 Arithmetic probes (E9)

Table 25 reports the accuracy of arithmetic comparisons by model.

Table 10: E1b (Scaled prices): estimates by condition.

Scale		Blocks	Slope	Naive	Left-digit	95% CI
10	prices x10	1	0.22	-1.25	-2.82*	[-2.82, -2.82]
100	prices x100	12	1.88	-7.12	-3.64*	[-5.95, -1.95]

Each condition is estimated on its own blocks. Slope is the within-block per-dollar demand response; naive is the uncorrected step-10 contrast; left-digit is step 10's residual from the placebo regression on destination roundness. * marks an interval excluding zero.

Table 11: E1b (Scaled prices): estimates by model.

Model	Blocks	Slope	Left-digit	95% CI
mistral-small-3.2-24b-instruct	13	1.75	-3.57*	[-5.32, -1.96]

B.11 STATED CONVENTION (E10)

Table 26 reports the ending effects under each preamble. Tables 27 and 28 report the left-digit estimates by condition and by model.

B.12 CORPUS FREQUENCY (E5)

Table 29 reports the posting frequency of each ending in three public corpora.

B.13 MATCHED DESIGN (E0)

Table 30 reports the estimates by condition. Table 31 gives the matched left-digit estimate, and Table 32 breaks this by model.

B.14 DIGIT POSITION (E0B)

The matched design (Section 6) establishes that changing the leftmost digit is associated with a negative response. Still, in two-digit prices this change necessarily occurs when the dollar figure crosses a ten. Three-digit prices allow us to separate these cases while holding the destination ending fixed at .00.

Consider three transitions, all ending at .00: \$214.99 → \$215.00, where no dollar digit changes (`flat`); \$219.99 → \$220.00, where the tens digit changes but the leading digit does not (`inner`); and \$199.99 → \$200.00, where the leading digit changes (`lead`). Let $\delta_{10}|_{\texttt{flat}}$, $\delta_{10}|_{\texttt{inner}}$, and $\delta_{10}|_{\texttt{lead}}$ denote the step-10 jump in each condition.

In the `flat` condition, purchase probability rises by 4.98 pp. In the `inner` condition, the jump is -0.91 pp. The difference,

$$\delta_{10}|_{\texttt{inner}} - \delta_{10}|_{\texttt{flat}} = -5.89 \quad [-8.77,\ -2.94],$$

measures the additional response associated with crossing the tens boundary. The observed jump in the `lead` condition is $\delta_{10}|_{\texttt{lead}} = -1.36$ pp. Relative to `flat`, the difference is

$$\delta_{10}|_{\texttt{lead}} - \delta_{10}|_{\texttt{flat}} = -6.34 \quad [-9.79,\ -3.11].$$

The excess relative to the extrapolated roundness gradient is:

$$\text{Excess} \;=\; \left(\delta_{10}|_{\texttt{lead}} - \delta_{10}|_{\texttt{flat}}\right) \;-\; 2\left(\delta_{10}|_{\texttt{inner}} - \delta_{10}|_{\texttt{flat}}\right) \tag{5}$$

The estimated excess is 5.44 pp, indicating that the leading-digit response is less negative than predicted by extrapolating the roundness gradient. In only 2 of 9 models is the excess negative.

Table 33 reports the estimates by condition. Table 34 gives the matched estimates, and Table 35 breaks this out by model.

Table 12: E2 (Prompt phrasing): estimates by condition.

Template	Blocks	Slope	Naive	Left-digit	95% CI
T01	75	-4.81	-6.37	-5.01*	[-9.79, -0.62]
T02	79	-3.61	-9.33	-10.33*	[-16.75, -4.48]
T03	68	-0.38	-14.88	-14.60*	[-21.84, -7.85]
T04	63	-0.77	-3.14	-10.67*	[-18.22, -2.93]
T05	76	-4.66	-7.03	-10.11*	[-18.89, -2.70]
T06	76	-2.98	-7.32	-6.24*	[-12.78, -0.45]
T07	71	-2.32	-7.04	-4.53	[-12.28, 6.35]
T08	79	-4.32	-7.46	-9.48*	[-15.29, -4.09]
T09	73	-2.25	-7.52	-6.46*	[-10.86, -0.79]
T10	71	-4.93	-14.56	-15.94*	[-24.35, -9.99]
T11	67	-2.47	-4.25	0.13	[-4.43, 5.36]
T12	74	-3.46	-12.97	-13.55*	[-20.15, -7.82]

Each condition is estimated on its own blocks. Slope is the within-block per-dollar demand response; naive is the uncorrected step-10 contrast; left-digit is step 10's residual from the placebo regression on destination roundness. * marks an interval excluding zero.

Table 13: E2 (Prompt phrasing): estimates by model.

Model	Blocks	Slope	Left-digit	95% CI
mistral-small-3.2-24b-instruct	270	-1.09	-11.82*	[-13.15, -10.44]
deepseek-v4-flash	246	-7.40	-8.50*	[-13.05, -3.76]
qwen3-235b-a22b-2507	174	0.98	-7.97*	[-11.73, -4.75]
gpt-4o-mini	152	-3.40	-6.54*	[-12.09, -0.64]
qwen3-30b-a3b-instruct-2507	30	-9.39	-3.49	[-14.89, 6.46]

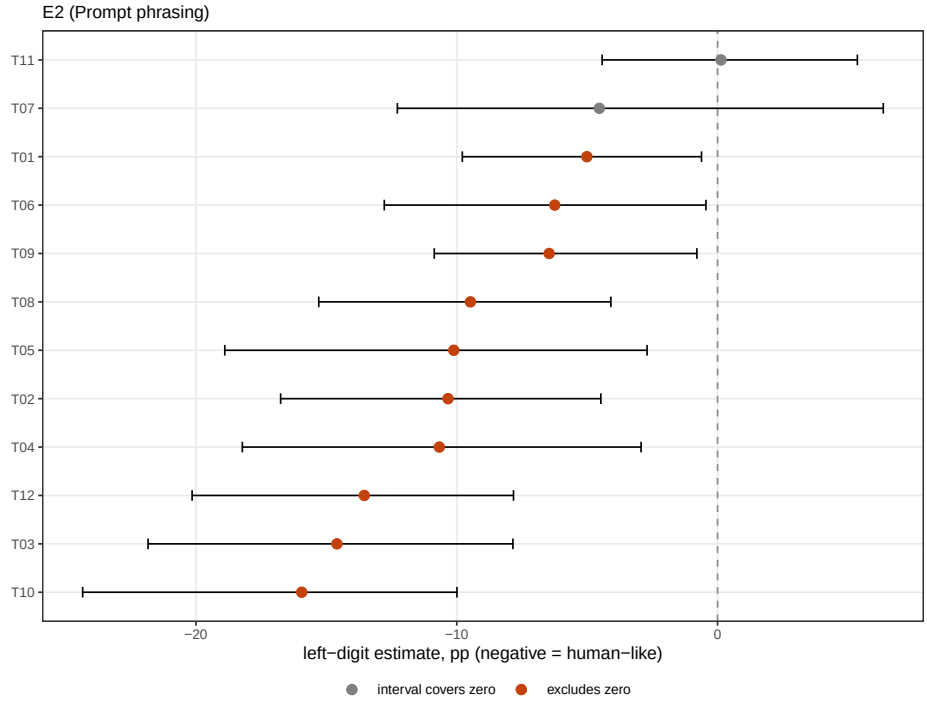

Figure 8: Estimates under each prompt template (E2).

Table 14: E3 (Non-price magnitudes): estimates by condition.

Domain		Blocks	Slope	Naive	Left-digit	95% CI
calories	calories per portion	893	-4.20	5.52	2.84*	[1.20, 4.43]
delivery_days	delivery, days	623	-4.64	9.67	6.18*	[4.22, 8.69]
odometer	odometer, miles	731	-2.93	-4.19	-4.97*	[-6.52, -3.21]
rating	rating out of 10	786	1.31	-0.47	-9.73*	[-13.16, -6.72]
weight	pack weight, kg	973	-6.12	5.73	-3.26*	[-5.49, -1.61]

Each condition is estimated on its own blocks. Slope is the within-block per-dollar demand response; naive is the uncorrected step-10 contrast; left-digit is step 10's residual from the placebo regression on destination roundness. * marks an interval excluding zero.

Table 15: E3 (Non-price magnitudes): estimates by model.

Model	Blocks	Slope	Left-digit	95% CI
gpt-4o-mini	729	-1.47	-11.37*	[-14.51, -8.25]
deepseek-v4-flash	789	-2.03	-1.66	[-3.78, 0.74]
qwen3-235b-a22b-2507	844	-4.65	-0.96	[-2.84, 1.22]
mistral-small-3.2-24b-instruct	959	-1.20	0.07	[-0.49, 0.65]
qwen3-30b-a3b-instruct-2507	685	-8.71	3.33*	[0.38, 6.39]

Table 16: E4 (Currency and locale): estimates by condition.

Currency		Blocks	Slope	Naive	Left-digit	95% CI
CHF	Swiss franc	98	-6.23	-3.43	-4.59	[-9.06, 0.26]
EUR	euro	82	-4.12	-3.20	-4.04	[-9.10, 1.27]
INR	Indian rupee	89	-6.49	16.22	3.82	[-3.09, 11.02]
JPY	Japanese yen	82	-0.92	4.13	-3.42	[-7.99, 1.71]
USD	US dollar	92	-5.07	-5.03	-5.28*	[-10.57, -0.05]

Each condition is estimated on its own blocks. Slope is the within-block per-dollar demand response; naive is the uncorrected step-10 contrast; left-digit is step 10's residual from the placebo regression on destination roundness. * marks an interval excluding zero.

Table 17: E4 (Currency and locale): estimates by model.

Model	Blocks	Slope	Left-digit	95% CI
deepseek-v4-flash	91	-5.71	-9.45*	[-14.89, -3.56]
qwen3-235b-a22b-2507	84	-0.73	-3.28	[-10.13, 3.56]
mistral-small-3.2-24b-instruct	115	-3.32	-1.11	[-2.81, 0.15]
gpt-4o-mini	72	-7.05	-0.39	[-7.80, 7.79]
qwen3-30b-a3b-instruct-2507	81	-7.38	1.03	[-4.39, 6.91]

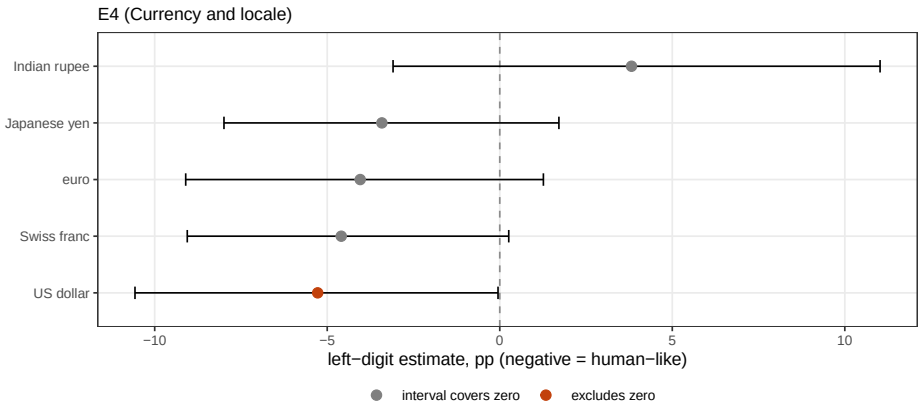

Figure 9: Estimates in each currency (E4).

Table 18: E6 (Number formatting): estimates by condition.

Rendering		Blocks	Slope	Naive	Left-digit	95% CI
digitspace	\$1 2 . 9 9	65	-4.18	-17.98	-3.11	[-12.07, 5.69]
iso	USD 12.99	69	-5.07	-0.27	0.59	[-5.04, 7.97]
plain	\$12.99	70	-5.14	-4.05	-2.68	[-10.52, 4.59]
spaced	\$ 12.99	69	-5.28	-2.70	-2.31	[-7.73, 5.42]
words	twelve dollars ninety-nine	67	-5.58	1.64	6.28*	[0.67, 12.66]

Each condition is estimated on its own blocks. Slope is the within-block per-dollar demand response; naive is the uncorrected step-10 contrast; left-digit is step 10's residual from the placebo regression on destination roundness. * marks an interval excluding zero.

Table 19: E6 (Number formatting): estimates by model.

Model	Blocks	Slope	Left-digit	95% CI
deepseek-v4-flash	84	-6.15	-6.26	[-11.46, 0.95]
qwen3-235b-a22b-2507	89	0.24	0.18	[-6.47, 5.28]
gpt-4o-mini	92	-4.22	0.72	[-5.10, 6.39]
qwen3-30b-a3b-instruct-2507	75	-11.16	4.74	[-1.76, 11.75]

Table 20: The vision ladder (E7).

Condition	Rendering	Blocks	Slope	Left-digit	95% CI
textonly	text only, no image	637	-3.08	-2.79*	[-4.23, -1.26]
same	cents at 100% height	667	-5.33	0.16	[-3.98, 4.68]
r70	cents at 70%	667	-3.72	-0.36	[-4.48, 3.97]
r50	cents at 50%	679	-5.68	-1.81	[-6.06, 1.94]
r35	cents at 35%	686	-5.77	-3.66	[-7.29, 0.83]
sup	cents as a superscript	55	-4.14	-18.66	[-41.68, 0.73]

Every condition posts the same prices; only the height of the cents relative to the dollar digits changes. `textonly` is the same task with no image at all.

Table 21: E7 (Rendered price tags): estimates by model.

Model	Blocks	Slope	Left-digit	95% CI
gpt-4o-mini	820	-8.27	-5.51*	[-9.88, -0.92]
qwen3.7-flash	1,358	-5.03	-3.45*	[-6.57, -0.64]
gpt-4o	56	-1.41	-1.30	[-4.87, 2.61]
mistral-small-3.2-24b-instruct	1,157	-2.03	2.24*	[1.61, 2.89]

Table 22: The cents-size gradient in E7, by model.

Model	Blocks	Trend	95% CI	$\Pr(\gamma > 0)$
mistral-small-3.2-24b-instruct	905	-0.21	[-2.93, 2.56]	0.440
gpt-4o-mini	681	25.53*	[0.88, 50.45]	0.975
qwen3.7-flash	1,079	-2.74	[-18.02, 11.90]	0.415
pooled	2,699	5.65	[-2.05, 14.95]	0.920

Trend is the slope of the left-digit estimate on the height of the cents relative to the dollar digits, in percentage points per unit of height. A positive value means the effect gets more negative as the cents shrink, which is what a perceptual account predicts. The last column is the share of bootstrap draws above zero. Models contributing fewer than 100 blocks are excluded.

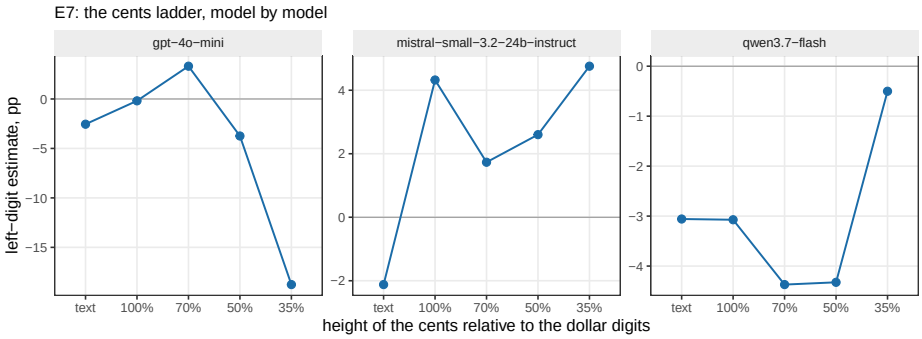

Figure 10: Effect of reducing the visual size of the cents. Each panel shows the left-digit estimate at four relative cent sizes.

Table 23: E8 (Explicit reasoning): estimates by condition.

Instruction		Blocks	Slope	Naive	Left-digit	95% CI
high	think step by step	10	-8.05	-9.12	-30.38	[-103.21, 9.28]
low	answer immediately	10	-0.90	17.18	15.03	[-30.57, 89.41]

Each condition is estimated on its own blocks. Slope is the within-block per-dollar demand response; naive is the uncorrected step-10 contrast; left-digit is step 10's residual from the placebo regression on destination roundness. * marks an interval excluding zero.

Table 24: E8 (Explicit reasoning): estimates by model.

Model	Blocks	Slope	Left-digit	95% CI
gpt-oss-20b	6	-14.33	-18.75	[-174.38, 99.58]
qwen3-235b-a22b-thinking-2507	6	1.31	-10.38	[-31.37, 10.40]
qwen3-235b-a22b-2507	3	0.07	0.19	[-0.42, 0.54]
deepseek-v4-flash	3	-3.67	3.17*	[0.74, 6.98]
deepseek-chat-v3-0324	2	-0.32	5.59	[-2.21, 13.39]

Table 25: Arithmetic probes (E9).

Model	Items	P(correct)	Median answer	Share saying 1
gemma-3-27b-it	62	1.000	1.00	1.000
gemma-4-31b-it	62	1.000	1.00	1.000
gpt-oss-120b	62	1.000	1.00	1.000
qwen3-14b	62	1.000	1.00	1.000
qwen3-32b	62	1.000	1.00	1.000
gpt-4o-mini	62	1.000	1.00	1.000
gpt-oss-20b	62	1.000	1.00	1.000
qwen3-30b-a3b-instruct-2507	62	1.000	1.00	1.000
gpt-4o	62	1.000	1.00	1.000
qwen3-235b-a22b-2507	62	1.000	1.00	1.000
deepseek-chat-v3-0324	62	1.000	1.00	1.000
mistral-small-3.2-24b-instruct	62	1.000	1.00	1.000
qwen3.7-flash	62	0.997	1.00	1.000
qwen3.5-9b	62	0.994	1.00	0.984
deepseek-v4-flash	62	0.978	1.00	0.871

Comparison items ask which of two prices is lower; magnitude items ask how many cents separate them.

Table 26: Installing a pricing convention (E10).

Preamble	Ending	Blocks	Effect	95% CI
none	.33	91	-1.46	[-3.65, 1.12]
p00	.33	100	-2.19*	[-4.10, -0.81]
p33	.33	107	36.02*	[29.79, 42.31]
p77	.33	102	1.65	[-0.35, 3.55]
none	.77	91	2.32*	[0.22, 4.72]
p00	.77	100	-2.43*	[-4.14, -0.96]
p33	.77	107	1.27*	[0.43, 2.37]
p77	.77	102	33.02*	[26.84, 38.23]

Effect is the ending's own effect minus the mean of the other endings in its decade, so a local price trend cannot generate it. The prediction is a positive entry exactly where the preamble names that ending.

Table 27: E10 (Stated convention): estimates by condition.

Preamble		Blocks	Slope	Naive	Left-digit	95% CI
none	no preamble	91	-4.79	-3.69	-0.87	[-7.16, 7.51]
p00	prices end in .00	100	0.89	26.67	28.71*	[21.73, 36.81]
p33	prices end in .33	107	-4.55	-1.04	1.36	[-3.64, 6.08]
p77	prices end in .77	102	4.81	-0.88	-4.41	[-10.58, 0.37]

Each condition is estimated on its own blocks. Slope is the within-block per-dollar demand response; naive is the uncorrected step-10 contrast; left-digit is step 10's residual from the placebo regression on destination roundness. * marks an interval excluding zero.

Table 28: E10 (Stated convention): estimates by model.

Model	Blocks	Slope	Left-digit	95% CI
mistral-small-3.2-24b-instruct	94	-0.78	1.56*	[0.56, 2.84]
deepseek-v4-flash	70	-2.52	3.02	[-1.37, 8.95]
qwen3-235b-a22b-2507	85	-0.98	3.42	[-5.37, 10.64]
gpt-4o-mini	64	-3.02	8.82*	[1.25, 17.02]
qwen3-30b-a3b-instruct-2507	87	2.10	14.64*	[4.50, 23.93]

Table 29: The twelve most common price endings on the open web (E5).

Ending	`c4`	`dolma`	`redpajama`	Mean
.99	33.40	30.35	36.91	33.55
.00	22.98	17.96	16.62	19.19
.95	17.65	15.07	14.95	15.89
.50	9.53	9.71	7.59	8.94
.25	3.20	3.55	4.08	3.61
.75	2.30	2.64	2.73	2.55
.90	1.41	1.53	1.91	1.61
.49	1.29	1.62	1.41	1.44
.20	1.07	1.32	0.99	1.12
.80	1.04	1.28	0.94	1.09
.60	1.02	1.27	0.94	1.08
.40	0.86	1.11	0.87	0.95

Percentage of all dollar-and-cents prices posted with each ending, counted with the infini-gram API over three public corpora. The ordering is nearly identical across the three (Spearman 0.968 to 0.989), which is why the corpus choice does not matter to the argument.

Table 30: E0 (Dollar boundary type): estimates by condition.

Dollar figure		Blocks	Slope	Naive	Left-digit	95% CI
roll	leftmost digit changes (19 to 20)	794	-1.65	-3.19	-5.08*	[-8.46, -1.59]
same	dollar figure increments (12 to 13)	840	-0.62	3.25	1.92	[-0.91, 4.55]
widen	a digit is added (99 to 100)	102	-1.04	-0.40	-2.37	[-7.34, 2.56]

Each condition is estimated on its own blocks. Slope is the within-block per-dollar demand response; naive is the uncorrected step-10 contrast; left-digit is step 10's residual from the placebo regression on destination roundness. * marks an interval excluding zero.

Table 31: The left-digit effect without a roundness adjustment (E0).

At the top of the block ...	Blocks	Step-10 jump	Against first row	95% CI
the dollar figure increments inside a ten (12 to 13)	840	3.27	reference	
the leftmost digit changes (19 to 20)	794	-2.87	-6.14*	[-8.36, -3.86]
the dollar figure gains a digit (99 to 100)	102	0.47	-2.81	[-6.63, 1.06]

Every block in E0 ends on .00, so the attractiveness of .00 is common to all three rows and cancels in the comparison. The estimate is a difference between two step-10 jumps and no roundness model enters it. Anchors are matched in level across conditions and every decade carries two products.

Table 32: The E0 contrast within each model.

Model	Blocks	Increments	Leftmost changes	Difference
qwen3-235b-a22b-2507	140	1.13	-23.54	-24.66
gpt-oss-20b	188	12.44	-5.86	-18.30
qwen3-30b-a3b-instruct-2507	103	-0.20	-7.40	-7.20
gpt-4o-mini	131	12.80	5.88	-6.91
deepseek-chat-v3-0324	192	0.67	-5.50	-6.18
qwen3.7-flash	297	1.74	-2.92	-4.67
deepseek-v4-flash	183	-0.28	-2.05	-1.78
qwen3.5-9b	281	0.30	1.20	0.89
mistral-small-3.2-24b-instruct	221	3.45	5.38	1.93

Blocks are the model's total across the three conditions.

Table 33: E0b (Digit position): estimates by condition.

Digit position		Blocks	Slope	Naive	Left-digit	95% CI
flat	no boundary (214 to 215)	414	1.11	5.40	7.48*	[4.23, 12.36]
inner	tens digit changes (219 to 220)	422	0.36	-0.93	-2.03	[-7.44, 2.38]
lead	leading digit changes (199 to 200)	431	0.15	-2.07	-4.89	[-10.74, 2.03]

Each condition is estimated on its own blocks. Slope is the within-block per-dollar demand response; naive is the uncorrected step-10 contrast; left-digit is step 10's residual from the placebo regression on destination roundness. * marks an interval excluding zero.

Table 34: Digit position versus roundness in three-digit prices (E0b).

At the top of the block ...	Blocks	Step-10 jump	Against flat	95% CI
no digit boundary ($214 \to 215$)	414	4.98	reference	
tens digit changes ($219 \to 220$)	422	-0.91	-5.89*	[-8.77, -2.94]
leading digit changes ($199 \to 200$)	431	-1.36	-6.34*	[-9.79, -3.11]

Three-digit prices separate digit position from roundness. Flat and inner differ only in destination roundness, with no leading-digit change in either; their gradient measures roundness alone. Lead adds the leading-digit change. If lead falls below the gradient extrapolated from flat and inner, the excess is a leading-digit effect.

Table 35: The E0b contrasts within each model.

Model	Blocks	Flat	Inner	Lead	Roundness	Lead excess
gpt-4o-mini	75	5.64	16.37	16.88	10.73	-10.21
mistral-small-3.2-24b-instruct	173	5.07	7.85	9.65	2.78	-0.98
deepseek-chat-v3-0324	148	6.73	4.12	2.34	-2.62	0.84
qwen3.5-9b	195	0.75	-1.06	0.89	-1.81	3.75
qwen3-235b-a22b-2507	104	3.42	-8.35	-15.70	-11.77	4.42
qwen3-30b-a3b-instruct-2507	55	-3.12	-1.80	4.95	1.32	5.44
qwen3.7-flash	211	4.70	-0.55	-0.27	-5.25	5.52
deepseek-v4-flash	157	4.99	-1.60	2.50	-6.59	10.70
gpt-oss-20b	149	12.71	-17.98	-28.77	-30.68	19.88

Roundness is inner minus flat; lead excess is lead minus the linear extrapolation from flat and inner.

Table 36: How much the endings differ by roundness, measured.

ρ	Destination is …	Endings	Mean effect	Corpus share, %
0	not a multiple of ten	20	-0.19	3.34
1	a plain multiple of ten	8	0.35	1.02
2	.50	1	0.37	8.94
3	a whole dollar	1	2.72	19.19

Mean effect is the average ending effect at that roundness level in the core sweep, in percentage points. Corpus share is the average posting frequency of endings at that level. The two orderings differ because .99 and .95 are among the most frequently posted endings but are not round, so the corpus-frequency adjustment uses distinct information from the ordinal score.

Table 37: The core estimate under three adjustments for destination roundness.

Adjustment for destination roundness	Estimate	95% CI
Ordinal roundness score (reported)	-3.25*	[-6.17, -0.65]
Measured corpus frequency of the destination	-3.52*	[-5.47, -1.66]
Largest placebo step (monotone bound)	-3.48*	[-5.40, -2.41]
Matched design, no roundness model (E0)	-6.14*	[-8.36, -3.86]

Row 1 is the reported estimator, which regresses the nine placebo jumps on an ordinal roundness score and extrapolates to a whole dollar. Row 2 replaces that score with the measured frequency of the destination ending in three public corpora, giving nine distinct values instead of two levels. Row 3 assumes only that rounder destinations are weakly more attractive: since .00 is the roundest destination on the grid, the dollar step should jump at least as much as the largest placebo, so the difference between them bounds the effect without any functional form. Row 4 uses no roundness model at all: it compares two step-10 jumps that both end on .00 and differ only in what happens to the dollar figure, and it is estimated on E0's own data. The first three rows are estimated on the core sweep.

C Robustness

C.1 Correction strategies

We use four approaches to separate the left-digit effect from destination roundness. The ordinal roundness correction and matched design are presented in the main text (Section 6). Two additional strategies are:

1. **Corpus frequency.** We replace the ordinal roundness score with the measured frequency with which each destination ending appears in web text. This allows the correction to use nine distinct empirical values instead of hard-coded discrete roundness levels.
2. **Monotone bound.** We make no functional-form assumption about roundness. We assume only that .00 is at least as attractive as the other destination endings. Under this assumption, the largest placebo jump provides a bound on the roundness component, and the remaining difference bounds the left-digit effect.

The four approaches produce similar estimates (Table 37). Table 36 reports the mean effect and corpus share at each roundness level.

C.2 The ten steps

Table 39 reports the ten one-cent steps from the core sweep.

C.3 Robustness of the estimator

Table 40 reports the left-digit estimate under alternative estimator choices.

Table 38: Left-digit effect by experiment, in percentage points of $P(\text{buy})$.

Experiment	Code	Blocks	Slope	Uncorrected	Left-digit	95% CI	θ
Dollar boundary type	E0	1,736	-1.11	0.09	-1.53	[-3.65, 0.57]	1.37
Digit position	E0b	1,267	0.53	0.75	0.11	[-2.77, 2.89]	n/a
Core sweep	E1	1,160	-4.28	-2.50	-3.25*	[-5.64, -0.99]	0.76
Scaled prices	E1b	13	1.75	-6.67	-3.57*	[-5.58, -1.85]	n/a
Newest models	E1f	48	-22.44	-1.60	-15.54	[-38.70, 9.16]	0.69
Prompt phrasing	E2	872	-3.15	-8.52	-8.91*	[-11.04, -6.93]	2.83
Non-price magnitudes	E3	4,006	-3.42	3.27	-2.01*	[-3.05, -0.93]	0.59
Currency and locale	E4	443	-4.67	1.63	-2.73*	[-5.04, -0.15]	0.58
Number formatting	E6	340	-5.06	-4.55	-0.26	[-3.42, 2.99]	0.05
Rendered price tags	E7	3,391	-4.73	2.01	-1.97*	[-3.69, -0.39]	0.42
Explicit reasoning	E8	20	-4.48	4.03	-7.68	[-56.93, 38.27]	1.71
Stated convention	E10	400	-0.86	5.33	6.22*	[2.92, 9.38]	n/a

Slope is the within-block per-dollar demand response. Uncorrected is the step-10 jump minus the mean of the nine placebo steps; it is confounded by the roundness of .00. Left-digit is step 10's residual from the placebo regression on destination roundness. Intervals are block bootstraps over 400 resamples; * marks intervals excluding zero. θ is the left-digit effect divided by the per-dollar slope, comparable to the inattention parameter in the human literature (Lyft ≈ 0.5, used cars 0.30, scanner data 0.15–0.25). Blank where the slope is within 1 pp of zero; n/a where the design does not produce a comparable parameter.

Table 39: The ten one-cent steps, core sweep (E1).

Step	Prices	Crosses $	Roundness	Jump	Roundness fit	Residual
1	.09 → .10	no	1	0.14	0.52	-0.38
2	.19 → .20	no	1	1.06	0.52	0.53
3	.29 → .30	no	1	0.18	0.52	-0.34
4	.39 → .40	no	1	1.54	0.52	1.02
5	.49 → .50	no	2	0.92	0.92	0.00
6	.59 → .60	no	1	1.16	0.52	0.63
7	.69 → .70	no	1	-1.30	0.52	-1.82
8	.79 → .80	no	1	0.42	0.52	-0.10
9	.89 → .90	no	1	0.99	0.52	0.46
10	.99 → +1.00	yes	3	-1.93	1.32	-3.25

Roundness codes the destination ending: 3 for .00, 2 for .50, 1 for any other multiple of ten, 0 otherwise. The fit is estimated on the nine placebo steps only; step 10's residual from it is the left-digit estimate.

Table 40: The left-digit estimate under each estimator choice.

Arm	Blocks	Linear	Quadratic	Cubic	Eight ρ=1	All blocks	Strict
E0	1,736	-1.53	-1.57	-1.56	0.18	-0.94	-2.06
E0b	1,267	0.11	0.08	0.12	0.79	0.07	0.16
E1	1,160	-3.25	-3.27	-3.10	-2.46	-2.10	-3.71
E1b	13	-3.57	-3.62	-3.80	-6.85	-0.20	-6.77
E1f	48	-15.54	-15.64	-15.03	-0.78	-6.22	-15.54
E2	872	-8.91	-8.88	-8.67	-8.50	-5.40	-10.84
E3	4,006	-2.01	-2.08	-1.88	3.58	-1.12	-2.69
E4	443	-2.73	-2.79	-2.51	1.89	-2.01	-3.30
E6	340	-0.26	-0.33	-0.10	-4.80	-0.19	-0.14
E7	3,391	-1.97	-2.04	-1.96	2.24	-1.49	-2.34
E8	20	-7.68	-7.63	-8.01	4.72	-5.11	-10.44
E10	400	6.22	6.06	6.33	5.27	5.18	7.24

Linear is the reported specification. Quadratic and cubic replace the within-block linear detrend with a polynomial of that degree. Eight ρ=1 drops the roundness extrapolation entirely and compares step 10 against the eight placebos landing on a plain multiple of ten. That leaves .00's extra roundness credited to the threshold, so it is the conservative direction: it understates the effect wherever the roundness correction matters, and the gap between the two columns is how much work the correction is doing. All blocks sets the minimum within-block range to zero, Strict to five percentage points.

Table 41: The core estimate with each model removed in turn.

Model dropped	Blocks remaining	Estimate
qwen3-30b-a3b-instruct-2507	1,102	-3.78
qwen3.5-9b	1,043	-3.66
gpt-oss-120b	1,104	-3.55
qwen3-32b	1,062	-3.50
mistral-small-3.2-24b-instruct	1,049	-3.46
qwen3-14b	1,077	-3.41
gpt-4o	1,095	-3.41
gemma-4-31b-it	1,140	-3.32
qwen3-235b-a22b-2507	1,077	-3.24
deepseek-chat-v3-0324	1,079	-3.20
gemma-3-27b-it	1,133	-3.18
gpt-4o-mini	1,091	-3.06
qwen3.7-flash	1,042	-3.05
deepseek-v4-flash	1,077	-2.95
gpt-oss-20b	1,069	-1.96

The full-sample estimate appears in the main table.

Table 42: Left-digit effect by condition within each arm (identification and prompt experiments).

Arm	Condition	Blocks	Slope	Left-digit	95% CI
E0	roll	794	-1.65	-5.08*	[-8.46, -1.59]
	same	840	-0.62	1.92	[-0.91, 4.55]
	widen	102	-1.04	-2.37	[-7.34, 2.56]
E0b	flat	414	1.11	7.48*	[4.23, 12.36]
	inner	422	0.36	-2.03	[-7.44, 2.38]
	lead	431	0.15	-4.89	[-10.74, 2.03]
E1b	10	1	0.22	-2.82*	[-2.82, -2.82]
	100	12	1.88	-3.64*	[-5.95, -1.95]
E2	T01	75	-4.81	-5.01*	[-9.79, -0.62]
	T02	79	-3.61	-10.33*	[-16.75, -4.48]
	T03	68	-0.38	-14.60*	[-21.84, -7.85]
	T04	63	-0.77	-10.67*	[-18.22, -2.93]
	T05	76	-4.66	-10.11*	[-18.89, -2.70]
	T06	76	-2.98	-6.24*	[-12.78, -0.45]
	T07	71	-2.32	-4.53	[-12.28, 6.35]
	T08	79	-4.32	-9.48*	[-15.29, -4.09]
	T09	73	-2.25	-6.46*	[-10.86, -0.79]
	T10	71	-4.93	-15.94*	[-24.35, -9.99]
	T11	67	-2.47	0.13	[-4.43, 5.36]
	T12	74	-3.46	-13.55*	[-20.15, -7.82]

Each arm's conditions are estimated on their own blocks, so the rows are independent rather than a decomposition of the arm total.

C.4 Leave one model out

Table 41 reports the pooled estimate after dropping each model in turn.

C.5 All per-condition estimates

Tables 42 and 43 collect every condition-level estimate across experiments.

Table 43: Left-digit effect by condition within each arm (domain, format, and extension experiments).

Arm	Condition	Blocks	Slope	Left-digit	95% CI
E3	calories	893	-4.20	2.84*	[1.20, 4.43]
	delivery_days	623	-4.64	6.18*	[4.22, 8.69]
	odometer	731	-2.93	-4.97*	[-6.52, -3.21]
	rating	786	1.31	-9.73*	[-13.16, -6.72]
	weight	973	-6.12	-3.26*	[-5.49, -1.61]
E4	CHF	98	-6.23	-4.59	[-9.06, 0.26]
	EUR	82	-4.12	-4.04	[-9.10, 1.27]
	INR	89	-6.49	3.82	[-3.09, 11.02]
	JPY	82	-0.92	-3.42	[-7.99, 1.71]
	USD	92	-5.07	-5.28*	[-10.57, -0.05]
E6	digitspace	65	-4.18	-3.11	[-12.07, 5.69]
	iso	69	-5.07	0.59	[-5.04, 7.97]
	plain	70	-5.14	-2.68	[-10.52, 4.59]
	spaced	69	-5.28	-2.31	[-7.73, 5.42]
	words	67	-5.58	6.28*	[0.67, 12.66]
E7	r35	686	-5.77	-3.66	[-7.29, 0.83]
	r50	679	-5.68	-1.81	[-6.06, 1.94]
	r70	667	-3.72	-0.36	[-4.48, 3.97]
	same	667	-5.33	0.16	[-3.98, 4.68]
	sup	55	-4.14	-18.66	[-41.68, 0.73]
	textonly	637	-3.08	-2.79*	[-4.23, -1.26]
E8	high	10	-8.05	-30.38	[-103.21, 9.28]
	low	10	-0.90	15.03	[-30.57, 89.41]
E10	none	91	-4.79	-0.87	[-7.16, 7.51]
	p00	100	0.89	28.71*	[21.73, 36.81]
	p33	107	-4.55	1.36	[-3.64, 6.08]
	p77	102	4.81	-4.41	[-10.58, 0.37]

Continued from Table 42. Each arm's conditions are estimated on their own blocks.

Table 44: Arm-level left-digit estimates with Benjamini–Hochberg correction.

Arm	95% CI	p	BH p
Dollar boundary type	[-3.65, 0.57]	0.16	0.24
Digit position	[-2.77, 2.89]	0.94	0.94
Core sweep	[-5.64, -0.99]	0.01*	0.03*
Scaled prices	[-5.58, -1.85]	$<.001$*	0.001*
Newest models	[-38.70, 9.16]	0.19	0.26
Prompt phrasing	[-11.04, -6.93]	$<.001$*	$<.001$*
Non-price magnitudes	[-3.05, -0.93]	$<.001$*	$<.001$*
Currency and locale	[-5.04, -0.15]	0.03*	0.05*
Number formatting	[-3.42, 2.99]	0.88	0.94
Rendered price tags	[-3.69, -0.39]	0.03*	0.05*
Explicit reasoning	[-56.93, 38.27]	0.75	0.89
Stated convention	[2.92, 9.38]	$<.001$*	$<.001$*

p-values are two-sided, computed from the bootstrap standard error under a normal approximation. The Benjamini–Hochberg (BH) correction is applied across the 12 arm-level tests. An asterisk marks $p < .05$.

D MULTIPLICITY CORRECTION

We report 143 confidence intervals across arm-level, condition-level, and model-level results. For multiplicity adjustment, we focus on the 12 arm-level tests of the left-digit effect, each testing whether the corresponding estimate differs from zero. Applying the Benjamini–Hochberg (BH) correction across these tests leaves 7 arms significant at the 5% level. The main finding that LLMs respond differently to prices just below versus just above a dollar boundary remains robust to adjustment for multiple testing.