Model-Directed Attention and the Persistence of Wrong Mental Models: Evidence from Firm Pricing*

Yi Han David Huffman Yiming Liu

September 16, 2026

Click to see the latest version

Abstract

We document a mechanism through which wrong mental models survive in data-rich environments: by remaining silent about the patterns they deem irrelevant, wrong models direct attention away from patterns that refute them. We examine this in firm pricing under consumer intertemporal substitution (IS), where consumers shift the timing of purchases, creating sales spikes on the discount day and dips before and after. A simple wrong model that low prices lead to high sales ignores IS, remains silent about the dips, and thus attributes the sales spikes entirely to new demand. Using survey and administrative data from 13,000 gas station managers, we find that over 40 percent neglect or underestimate intertemporal substitution, regardless of experience. Managers who hold the simple model perceive fuel demand to be more elastic, set lower prices, and earn lower profits. Online experiments with Prolific subjects provide causal evidence: holding the data fixed, randomly assigning the wrong model rather than the correct one reduces attention to the dips and increases neglect of intertemporal substitution, while prompting attention to the dips restores noticing and reduces neglect. Encountering the correct model later dislodges the wrong one, though not completely, while encountering the wrong one after the correct model has no effect.

*Han: Renmin University of China; yihanecon@ruc.edu.cn. Huffman: Cornell University; dbh234@Cornell.edu. Liu: Humboldt University of Berlin and WZB; yiming.liu@hu-berlin.de.

1 Introduction

Traditional economic theory assumes decision makers hold a correct model of which factors determine the outcomes of interest, and rationally update beliefs about model parameters as data arrive, converging to the truth in a data-rich environment. An important advance has been theories which weaken the first half of this assumption, and explore how beliefs need not converge to the truth, no matter how much data arrive, if the model itself is wrong (Berk, 1966; Esponda and Pouzo, 2016; Spiegler, 2016; Heidhues et al., 2018); see Bohren and Hauser (2025) for a review. This has led in turn to a growing literature on how a decision who holds a wrong "mental model" can fail to abandon it (e.g., Hong et al., 2007; Ortoleva, 2012; Schwartzstein and Sunderam, 2021; Fudenberg and Lanzani, 2023).

This paper provides the first direct test of model-directed attention, a mechanism that can lead a decision maker to keep a wrong mental model even when the disconfirming data are in front of them. Formalized by Gagnon-Bartsch et al. (2026), the idea is that a mental model directs attention towards the patterns it explains. A broad class of overly simple wrong models can therefore persist: those that fit well the patterns they explain but are silent about the rest direct attention away from precisely the patterns that would disconfirm them. A second implication concerns encountering another model. The wrong model can be dislodged by the correct one, but only if the correct model comes to direct attention towards the features the wrong model cannot explain; the correct model should not be dislodged by the wrong one, since those features have already been noticed. We test both implications in two settings. In the field we study roughly 13,000 gas station managers, who see ample data daily and can determine prices; online, we run experiments with Prolific subjects. These designs let us do what is otherwise difficult: hold the data fixed in an environment where we know the correct model, assign the initial mental model exogenously, measure what subjects notice pattern by pattern, and then either manipulate attention separately from the model or challenge that model with a second one.

Our investigation focuses on a wrong model that is about one of the most fundamental causal relationships in market settings: The effect of a price change on consumer behavior. We hypothesize that some decision makers operate under a wrong model in which the increase in sales that accompanies a price cut is attributed entirely to increased demand. The model says nothing about a second channel: regular customers shifting the timing of purchases they would have made anyway. Intertemporal substitution (IS) leaves a distinctive signature in the data: sales dip immediately before and immediately after the pre-announced discount. Thus the evidence required to refute the wrong model is present in the sales data decision makers see. Because the wrong model neglects IS, it gives no reason to notice differences across sales for regular-price days, and thus leads to not noticing the dips that would refute it.

Section 2 of the paper introduces our main measure of whether a decision maker has a

mental model of consumer behavior that neglects IS, and provides a conceptual framework for how this can persist due to model-directed attention. Our measure involves presenting subjects with hypothetical data on temporary price cuts by an online detergent seller, where subjects know we created the data using underlying rules but are not told what they are. We choose detergent because of previous evidence that consumers engage in IS in detergent purchasing (Hendel and Nevo, 2006b). The detergent task shows two repetitions of the exact same pattern in a table format, with zero noise: High sales during discounts but low sales immediately before and after, reflecting some consumers shifting purchase timing (it also shows that competitor price is constant). With the detergent task we know all participants have the same data in front of them, and there is an objectively correct interpretation. After seeing the data, subjects predict sales under permanent low pricing. Predicting the same value of sales under permanent low pricing as shown in the data for discount days (a value of 202) is our main indicator of neglecting IS.

Our conceptual framework shows how the simple wrong model leads to not noticing the dips in the detergent data, and fits the data as well as the correct model based on what is noticed. Given the wrong model persists due to directed attention, our model also implies that directing attention to the dips should restore attention and dislodge the wrong model. We derive further implications, showing that neglect of IS should lead to believing demand is more elastic than holding the correct model, and to thinking the optimal price is lower. Finally, we derive an asymmetric effect of encountering a second mental model: The correct model is not given up after encountering the wrong model, but the wrong model can be given up after encountering the correct model.

Section 3 of the paper provides evidence from the field, beginning with a description of the field environment we consider and the datasets we use. We leverage a collaboration with a firm that operates more than 20,000 gas stations. Each station is operated by a manager who has substantial influence on setting fuel prices at their station. In a survey sent to all of the station managers, we presented managers with the detergent task. We also elicited manager beliefs about demand elasticities for fuel and beliefs about the profitability of price cuts for fuel. We can match these measures to panel data on the managers' actual pricing decisions in their stations. Transaction-level data provide evidence that consumers do, in fact, exhibit substantial intertemporal substitution in response to fuel price discounts in the markets we study.

We find that more than 20 percent of managers completely neglect intertemporal substitution in their predictions of consumer behavior based on the detergent data, with another 19 percent underestimating. This shows persistence of the wrong mental model in the face of two cycles of zero-noise data that refute it. Neglecting intertemporal substitution in the detergent task is only weakly negatively correlated with manager experience, indicating the persistence of this blind spot. We also find evidence consistent with the hypothesized mechanism of model-directed attention: Managers who hold a mental model that lacks IS, are

significantly less likely to mention the pre-discount dip and the post-discount dip in open-text explanations of the data, consistent with not having noticed these. Finally, we show that IS neglect has important consequences: (1) explaining how managers interpret data on other types of price cuts in real gas station data; (2) leading to beliefs that fuel demand is more elastic; (3) charging lower average prices for fuel; (4) having 3.3% lower monthly profits through the channel of charging lower prices.

Section 4 tests the mechanism directly, using two sets of online experiments (Study 1 and Study 2) with an independent sample from Prolific, and continuing with IS-neglect as our workhorse example. The theory is largely silent about where initial mental models come from; our approach to testing the impact of mental models on attention and beliefs is to exogenously vary which peer mental model subjects are assigned. This mirrors evidence from our field setting involving gas station managers, where many managers report learning from mentors or peers, primarily early in their careers. To elicit peer mental models, we conducted an initial pilot in which we showed subjects the same type of artificial data about detergent discounts and sales as the station managers, and elicited their explanations.

Study 1 tests how a wrong model persists against disconfirming data. Subjects are assigned either a correct model involving IS or a wrong model in which discounts simply raise sales. Being assigned the wrong model reduces attention to the dips: the modal subject in that treatment does not recall that sales vary across regular-price days, either failing to recall the dip-day values or reporting them as equal to an unaffected regular day. It also raises neglect of IS in the incentivized detergent task. A third treatment adds an attention prompt to the wrong model, asking subjects to explain why sales are lower on the dip days than on other regular-price days. The prompt restores recall of the dip days and significantly lowers IS-neglect relative to the wrong-model treatment, establishing a causal effect of attention to the disconfirming patterns. Together with the first finding, this closes the loop: the model directs attention, and redirected attention revises the model.

Study 2 asks whether encountering a peer's model dislodges the one a subject already holds. The prediction is asymmetric, so we test both orders: subjects are assigned the wrong or the correct model in stage 1, make a prediction as in Study 1, and then in stage 2 encounter the other model. Encountering the wrong model should do little to those who already hold the correct one, who may not entertain it or, if they do, find that it cannot explain what they have already noticed. Encountering the correct model may move those who hold the wrong one, since it directs attention to the patterns their model deems irrelevant. Only the pair of orders separates this from a general tendency to adopt whichever model comes second. We find almost no change when the correct model comes first, and strong movement off the wrong model when it comes second. The movement is incomplete: IS-neglect in stage 2 remains significantly higher among subjects assigned the wrong model first. This is consistent with some subjects not entertaining the second model and continuing to read the data through the first, and implies that encountering the correct model may not entirely

undo the influence of having held a wrong one.

Section 5 concludes with a focus on the broader implications of our findings. The mechanism we demonstrate is a way for wrong models to persist under the seemingly unforgiving condition of exogenously arriving data that refute them, and it allows the problematic models to be identified ex ante. We argue that such simple, wrong models are likely to be pervasive and to lead to large mistakes, and we discuss examples from a range of settings.

We inform a theoretical literature on why a wrong model is not revised. One explanation concerns situations where data are generated endogenously by actions, and the resulting data fits the wrong model best (Cho and Kasa, 2015; Heidhues et al., 2021; Fudenberg et al., 2021; Ba, 2026). A second is that a misspecification can survive in evolutionary terms if it induces actions with better payoffs than the correct model (Fudenberg and Lanzani, 2023). A third holds that simple models may be retained because more complex ones are costly to adopt (Hong et al., 2007; Kendall and Oprea, 2024). A fourth is that a wrong model can persist because the correct model is outside the consideration set (Ortoleva, 2012; Schwartzstein and Sunderam, 2021). Our findings show a role for model-directed attention as a distinct source of persistent wrong models and beliefs. In the setting of our design, the first three mechanisms are excluded: disconfirming data arrive exogenously; the wrong model generates lower payoffs, for both managers and Prolific subjects; and our attention prompt does not reduce complexity, yet help dislodge the wrong model. The fourth account we do not rule out by design, but it cannot explain two of our main findings: Our attention prompts supply no alternative model, so a consideration-set account predicts no effect from them, yet they dislodge the wrong model; placing the correct model directly into the consideration set does not fully correct beliefs, with some subjects retaining the wrong model after the correct one is supplied.

Our results are complementary to another theoretical literature on how decision makers who understand their environment correctly can have biased beliefs. Under rational inattention, a decision maker who knows the structure of the problem trades precision against the cost of attention (Sims, 2003; Maćkowiak et al., 2023). Under costly cognition, one may know which variables matter yet neglect those whose payoff consequences are small relative to the cost of tracking them (Gabaix, 2014, 2019), or shade answers toward priors when uncertain about one's own reasoning (Enke and Graeber, 2023). Our results on the source of wrong beliefs are not well explained by rational inatttention or cognitive costs. IS-neglect costs managers about 4 percent of monthly profits and many managers do account for it, so it is neither trivial nor beyond reach. In the experiments, data, task and incentives are fixed and only the assigned model varies, so no such calculus should differ across treatments; and though dips and spike are equally legible cells in the same table, only the dips are recalled worse, and only under the wrong model. A version of rational inattention with a prior over which dimensions are informative can match this, but only by letting that prior govern attention—the mechanism we test, under another name. Motivated reasoning (Bénabou

and Tirole, 2016; Zimmermann, 2020; Golman et al., 2017) could produce bias of the kind we document, if decision makers prefer to believe customers respond as the simple model implies. But it is unclear why managers would prefer that to the profits accurate beliefs would bring, and Prolific subjects have still less stake, are paid for accuracy, and should not have a change in motive based on a randomly assigned peer's explanation.

We also contribute to the laboratory and online experimental literature on misspecified and competing mental models (for a review, see Ambuehl et al., 2026). Fit appears to govern adoption: narratives that explain the data better persuade more (Barron and Fries, 2024), and many subjects reject a model contradicted by the data when a rival is shown alongside it (Ambuehl and Thysen, 2025). Yet wrong models persist against data that should dislodge them—randomly assigned false narratives distort choices even though subjects see everything (Charles and Kendall, 2025)—and in the starkest case, subjects given the parameters of an updating problem begin with base-rate neglect and, after 200 rounds of feedback, hold beliefs further from the Bayesian benchmark than subjects not given the parameters (Esponda et al., 2024). Gagnon-Bartsch et al. (2026) read this through their framework: removing uncertainty about parameters a decision maker would eventually rule out can hinder discovery of an error, by reducing the perceived need for attention.[1] We test model-directed attention directly, in three respects. First, the mental model is our treatment, assigned at random with the data held fixed, as Charles and Kendall (2025) do for narratives. Second, we measure attention pattern by pattern, through incentivized recall after the data are removed, since the mechanism predicts that the wrong model lowers recall of the patterns it deems irrelevant but not of those it explains.[2] Third, we manipulate attention while holding the assigned model fixed, using prompts that point to the patterns the wrong model deems irrelevant.

We also complement field evidence on misspecified mental models. Households hold heterogeneous mental models of the macroeconomy and of inflation (Andre et al., 2022, 2026a), and most households and some financial professionals expect stale news about a company's earnings to predict its future stock returns (Andre et al., 2026b). In the same

[1]Related experiments show that models formed from data are imperfect and sticky: some subjects ignore relevant variables while others use irrelevant ones (Fréchette et al., 2025), subjects converge on the same simple models when several fit (Kendall and Oprea, 2024), a model formed before a new variable was seen continues to influence beliefs after revision (Grass et al., 2026), and many stop updating correctly once parameters change (Esponda et al., 2023). A second group documents errors when the information needed to avoid them is displayed: neglect of endogenous sample selection (Esponda and Vespa, 2018), of what a transparent selection process withheld (Enke, 2020), of a second, payoff-irrelevant cause of a signal (Graeber, 2023), and of correlation among signals (Enke and Zimmermann, 2019). In these latter designs the experimenter chooses which feature of the data will go unnoticed and then documents that it does; in ours, the model is randomized and what goes unnoticed follows from it.

[2]Prior work measures attention more coarsely: how many rounds of feedback subjects choose to view (Esponda et al., 2024), which charts they open when rival models disagree (Ambuehl and Thysen, 2025), and how a narrative shifts the weight placed on different years of data (Barron and Fries, 2024). Closest to ours, an online appendix of Esponda et al. (2024) elicits incentivized recall of signal-outcome frequencies and finds that subjects who begin with base-rate neglect recall them with larger errors, in the direction of their beliefs.

setting we study, Han et al. (2025a) show that managers with lower cognitive skills underappreciate how competitors respond to price cuts. These studies elicit the models people hold through hypothetical scenarios, open-ended explanations, and direct belief questions, and trace where they come from, as when varying news consumption shows the media to be an important source of macroeconomic narratives (Andre et al., 2026a). Our study is different in examining how a mental model shapes what a decision maker notices in a given body of data. Closest to our mechanism, Hanna et al. (2014) show that seaweed farmers fail to optimize pod size even though their own plots generate data on its value, and change their choices only after receiving a summary that highlights the relationship and recommends a size. There too attention is inferred indirectly, and the intervention bundles three things at once: it directs attention to a dimension, supplies the relationship that explains it, and states the answer. We separate these by measuring the mental model directly and attention pattern by pattern, and we showing that directing attention alone, with no relationship stated and no recommendation given, dislodges the wrong model without installing the right one, while supplying the correct model does more, though not everything.

Our paper has implications for a literature on behavioral IO. Many studies examine behavioral biases of consumers and how firms rationally exploit them (for a survey, see Heidhues and Kőszegi 2018), but there has been less work on potential biases of managers about consumers. A notable exception is Strulov-Shlain (2023), who finds that firms' pricing decisions are consistent with partial unawareness of consumers' left-digit bias (see also List et al. 2023).[3] We add evidence, from directly eliciting managers' mental models in surveys, that some firm managers neglect a basic feature of consumer behavior, intertemporal substitution. Neglecting it is associated with believing that demand is more elastic and with a greater tendency to view price cuts as good for profits. We also provide evidence on psychological mechanisms that can explain why this mistake persists and why managers differ in it.

Our findings also bear on a literature that asks why firms offer periodic, temporary price discounts. Studies of many product markets find that some consumers shift purchases to discount periods, presumably in part because they are more price sensitive (e.g., Boizot et al. 2001; Pesendorfer 2002; Hendel and Nevo 2006a, 2006b). These findings have led to an explanation of discounts as dynamic price discrimination. Our evidence on how managers think about intertemporal substitution shows that some managers do not underestimate it and may use discounts to price discriminate. Other managers underestimate it and favor discounts even more, consistent with overestimating how effective discounts are at attracting new customers. At least some discounts may therefore reflect managers' systematic misperception of consumer behavior rather than dynamic price discrimination.

[3]A nascent literature on "behavioral firms" documents mistakes in firm behavior and manager beliefs (e.g., Hortaçsu and Puller 2008; List and Mason 2011; Goldfarb and Xiao 2011, 2024; Hanna et al. 2014; DellaVigna and Gentzkow 2019; Hortaçsu et al. 2019; Tadelis et al. 2023). For a survey of an older tradition of bounded rationality in IO, see Ellison (2006).

2 Conceptual Framework

This section provides a conceptual framework for how model-directed attention may sustain an overly simple model of consumer behavior that neglects IS. We explain how a manager's mental model of demand directs his attention, why a wrong model can persist despite unlimited exposure to data that it fits worse than the correct model, and what dislodges it. The framework adapts the theory of channeled attention of Gagnon-Bartsch, Rabin and Schwartzstein (2026) to a pricing environment. Proofs are given in Appendix A.

The data the manager sees. A manager with a certain model of demand sees the price p_t and sales q_t on day t. There is a regular price, and a discounted price occurring on one day of each week. The timing of this discount is exogenously set and does not respond to demand. It is announced in advance so that customers can anticipate. A day is a discount day when its price is below the price on the days around it, and d_t is a day's signed distance to the nearest discount day, with ties broken arbitrarily. Sales are generated by

$$q_t \;=\; D^*(p_t) + \delta^*(d_t), \tag{1}$$

where $D^*(p)$ is demand at a permanent price p and δ^* is intertemporal substitution: an anticipated discount shifts purchases toward the discount day from the days around it. Throughout, intertemporal substitution is zero outside $d \in \{-1, 0, 1, 2\}$ and sums to zero; it reflects the pure change in timing of purchasing without any new demand.

Table 1: Laundry detergent sales data presented to managers

Day	Retailer A's price	Competitor B's price	Retailer A's sales volume	Retailer A's profit
Day 1	5.9	5.9	100	160
Day 2	5.9	5.9	100	160
Day 3	5.9	5.9	100	160
Day 4	5.9	5.9	90	144
Day 5	5.2	5.9	202	182
Day 6	5.9	5.9	82	131
Day 7	5.9	5.9	91	146
Day 8	5.9	5.9	100	160
Day 9	5.9	5.9	100	160
Day 10	5.9	5.9	90	144
Day 11	5.2	5.9	202	182
Day 12	5.9	5.9	82	131
Day 13	5.9	5.9	91	146
Day 14	5.9	5.9	100	160

Notes: This table shows the data managers see in the survey. The data generating process (DGP) involves both genuine demand response to prices and intertemporal substitution as specified in Equation 1. Managers are told that the data are representative and cover all possible situations, and that the table shows all the variables that may affect sales and profits. Managers are also informed that the marginal cost is 4.3.

Table 1 presents the data generated by Equation 1, where $D^*(5.9) = 100$ and $D^*(5.2) =$

165, and δ^* equals -10, 37, -18 and -9 at distances -1, 0, 1 and 2: sales are 100 on baseline days, dip to 90, 82 and 91 around each discount, and reach 202 on discount days. The table covers two weeks, but the rule is deterministic and thus the data is representative. A manager with the correct mental model of the data-generating process is able to work out the demand curve.

The simple wrong model and the correct model. A mental model is a subjective causal structure of the data generating process (DGP). In our pricing application, the mental model of a manager involves both what factors matter for sales, and how they interact to determine sales. Based on the mental model, the manager forms a belief about elasticity of demand. The *correct model* has the same structure as the true data-generating process, $q_t = D(p_t) + \delta(d_t)$, with D and δ to be learned through data. Its holder treats a day's position relative to the discount as informative, as the distance determines the amount of intertemporal substitution. Thus the high sales, 202, on the discount day is partly new demand, and partly borrowed from the surrounding days: $\delta(0) = -(\delta(-1)+\delta(1)+\delta(2)) = -(-10-18-9) = 37$. The demand at price 5.2, $D(5.2)$, can be backed out from $D(5.2) = 202 - \delta(0) = 165$. Baseline days give $D(5.9) = 100$, and the elasticity of demand equals $\frac{(D(5.9)-D(5.2))/D(5.9)}{(5.9-5.2)/5.9} = -5.48$.

The *simple wrong model* omits intertemporal substitution: $q_t = D(p_t)$. According to this model, price is the only factor that determines sales. Given the environment is deterministic, this model implies that each price has one sales value. The only thing manager of this model needs to learn is the value at each price. Thus when looking at the table, he only needs to record one value at each price, and his model then tells him no further day can teach him anything, because every regular-price day, and every discount day, is the same day. Learning is symmetric across the two prices and, given the model, correct. The mistake is the structure: a model allowing one value at each price rules out 90, 82 and 91, which the true DGP produces at the regular price. In the taxonomy of Gagnon-Bartsch, Rabin and Schwartzstein (2026), once he has read the table, the simple wrong model is *censored*. The two models therefore explain different patterns. The correct model explains the differences in sales among regular-price days by each day's position relative to the discount. The simple wrong model explains only the difference in sales between the two prices: under it all regular-price days fall in one cell of a partition, and $q_t \in \{82, 90, 91\}$ are censored. Thus, according to this model, baseline days give $D(5.9) = 100$ and the discount days give $D(5.9) = 202$, and the elasticity of demand equals $\frac{(D(5.9)-D(5.2))/D(5.9)}{(5.9-5.2)/5.9} = -8.60$.

Assumption 1 (Model-directed noticing)**.** *The manager's model directs his attention towards the patterns it explains: he notices a distinction in the data if and only if his model allows sales to differ across it. He gives up his model if and only if the data* as he has noticed them *are impossible under the model, or become infinitely less likely under it than under the correct model.*

The first part of the assumption has the model direct attention towards the patterns it explains and away from those that would refute it. The second part makes the manager judge a model by its fit to the noticed data, not by the full data on display, and we adopt the assumption of Gagnon-Bartsch, Rabin and Schwartzstein (2026) that the alternative model to compare with is the true model. In our deterministic environment, this means that the alternative model always fully explain the data. Thus for a wrong model to persist, it has to be that the wrong model also fully explain the data that it directs attention to. This assumption makes it harder for a wrong model to persist, and our main results hold if we instead assume that the alternative model is not the correct model and do not fully fit the data.

Why the simple wrong model persists despite the data. The simple wrong model prescribes a simple noticing strategy: read one value at each price and notice nothing further about sales, since every other day falls at a price already known. The manager sees the dips but does not notice them as values, and his model gives him no reason to assemble the comparison that would expose the contradiction, though the data table is in full view.

Prediction 1 (Attentional stability)**.** *Under Assumption 1 the simple wrong model is attentionally stable: following any noticing strategy the assumption prescribes, the manager does not give up the model, however much data accumulates.*

Given its noticing strategy, the simple wrong model never fits the noticed data more than a fixed factor worse than the correct model does (Appendix A.1). The data in front of the manager refute it, since 82 beside 100 at the same price is impossible under it, but the noticed data contain no such pair. It fully explains the variation it deems relevant: lower prices lead to higher sales when comparing the sales on the discount day to sales on the regular day.

Perceived elasticity and price. Holding the wrong model has direct implications for the belief about demand elasticity and the optimal price. The two models interpret the high sales on the discount day, 202, differently. While managers who hold the correct mental model realize that 37 out of 202 is borrowed from the surrounding days through intertemporal substitution, managers who hold the wrong model interpret the whole high sales as new demand. Thus, they form different beliefs about elasticity seeing the same data, and those who hold the wrong model perceive the demand to be more elastic.

Prediction 2 (Elasticity and prices)**.** *A holder of the simple wrong model perceives demand as more elastic than it is. Assuming he otherwise prices optimally, he sets lower prices than those of the correct model, and obtain lower profits.*

Correcting the model by directing attention. The stability in Prediction 1 is the stability of a noticing strategy, which identifies the remedy: impose the noticing the model declines

to do. Consider an attention prompt that direct managers' attention to the patterns they see on the pre-discount dip (e.g. Day 3 to Day 4) and the post-discount dip (e.g. Day 6 to Day 8). This attention prompt should force managers to notice that sales are not always 100 when prices are high, and there are more than one value under the same price. It is minimum: it supplies no information that is not already on the page, provides no alternative model, and no suggestion that position relative to the discount is what matters. However, given the stability is established through censoring certain values, this minimum intervention may dislodge the wrong model.

Prediction 3 (Directed noticing)**.** *Any noticing strategy that separately notices the sales values of two regular-price days with different sales yields noticed data with probability zero under the simple wrong model. Directing attention to the dip days therefore dislodges it.*

Once a noticed 82 stands beside a noticed 100 at the same price, no candidate of the simple wrong model accommodates the pair, and the manager gives up the simple wrong model: it now fits the noticed data infinitely worse than any model that admits the pair, the correct model among them.

Encountering the correct model. A peer or a mentor who articulates how the world works can also dislodge a model, but only if the alternative is *entertained*: the manager grants it some provisional weight α, if only as a passing thought, which extends Assumption 1 to it. An alternative described but not entertained changes nothing, because the patterns it describes are ones the manager has already seen, and through the lens of the incumbent model they mean nothing. He dismisses the model along with them.

Prediction 4 (Asymmetric correction)**.** *Consider a manager who sees data through the lens of one model encounters a second model, and he entertains the second model with probability $\alpha \in (0, 1)$. (i) A holder of the simple wrong model who entertains the correct model gives up the simple wrong model for the correct model. (ii) A holder of the correct model who entertains the simple wrong model continue to hold the correct model.*

The correct model conceptualizes everything the simple wrong model does and more. Data noticed under it refute the simple wrong model, while data noticed under the simple wrong model are too coarse to refute anything. The framework does not deliver α: whether an alternative is entertained lies outside it, and the framework gives the direction of the asymmetry but not the size of the correction. We treat entertaining as an empirical question, with confidence in the incumbent model as an inverse proxy.

3 Evidence from the field

3.1 Institutional context: Market structure, datasets, managers, relevance of intertemporal substitution in fuel markets

Our partner company operates more than 20,000 gas stations across the country. Stations typically sell both gasoline and diesel fuel, and essentially always have a convenience store. In this country and company, the bulk of station profits come from fuel sales rather than the convenience store; on average in our dataset, fuel profits make up 71% of total station profits. The stations are primarily company-owned, rather than franchises. Each station has a station manager who has substantial influence over station operations, including pricing decisions. Station operation is also governed, however, by the policies of more senior, district-level managers. There are about 350 districts, each with a district-level manager who sets policies about precisely what type and degree of discretion is given to station managers operating in their district.

The market is relatively concentrated, with important types of brand differentiation. Our partner company is one of two major brands in the market for retail gasoline, with stations all across the country and each claiming about one third of the market. The rest of the market share goes to hundreds of smaller, more local chains, which we refer to as "independent" competitors. One key difference between the major brands and independent brands is that the former produce their own fuel, while the latter must buy fuel products on the market. Another key difference is that the larger companies position themselves as offering a premium fuel product, and for this reason typically charge more than the independent companies for the same grade of gasoline or diesel.

An important feature of the pricing environment is that there is a government-imposed price ceiling for fuel products, indexed to the world price of oil. The price ceiling arguably serves as a natural focal point for coordinating pricing, and indeed, the two large companies have a policy of generally pricing near the price ceiling. Independent stations also frequently exhibit pricing that tracks movements in the price ceiling, albeit with a substantial discount. We measure a station's pricing policy by its price ratio, which our partner company computes as the sales-weighted average of each fuel product's price relative to its price ceiling. A price ratio of 1 means that the station prices at the ceiling for all products. The price ratio includes all discounts, such as posted-price cuts, coupons and promotions.

3.1.1 Datasets

We obtained data through a collaboration with the partner company's research arm.[4] This includes performance data (e.g., profits and prices), but also survey data collected via the

[4]IRB approval for this study was granted by Renmin University of China.

company's internal survey infrastructure. Because participation in internal surveys is expected—but these are administered by a unit independent of senior management—response rates are high and confidentiality is credible. Data from the human resources department of the company were not available. For this reason, our surveys were designed to collect variables that describe some aspects of the work environment that would normally be collected by human resources, such as manager work history. We were not allowed to collect certain variables, however, such as manager earnings or work hours.

Our main analysis is based on three datasets. The first is from a survey conducted with senior, district-level managers. We have responses from 353, close to a 100 percent response rate. One purpose of this survey was to collect systematic information about the amount of discretion given to station managers over fuel pricing. Another was to elicit senior manager views on potential mistakes by station managers when it comes to pricing.

The second dataset comes from a comprehensive survey conducted with station managers. The survey was sent to all 20,000 station managers, with a response rate of approximately 65 percent, yielding about 13,000 responses. This survey was designed to measure managers' neglect of intertemporal substitution through multiple approaches: the abstract detergent task, predictions about fuel demand elasticity at both typical and own stations, and interpretation of real historical pricing data. The survey also collected detailed information about manager characteristics, work history, and station operations.

The third dataset consists of monthly performance data on the company's gas stations for the period 2019 to 2024. These are panel data for each station, recording key outcomes such as fuel and nonfuel profits, sales volume, and average monthly prices charged for gasoline and diesel products. We have access to the monthly performance data in 26 of the 31 regions the company operates.[5] The total dataset covers approximately 16,000 gas stations. The research arm matched performance data to survey responses using an internal company identifier, yielding linked data for roughly 10,000 stations (due to the 65 percent survey response rate).

This combination of survey and administrative data allows us to link managers' mental models of customer behavior, as measured through their survey responses, to actual pricing decisions and station performance over a six-year period.

3.1.2 Manager characteristics, discretion, incentives

Table 2 shows descriptive statistics for station managers and their stations. The median age of a station manager is 39, and about 34 percent of managers are female. The modal level of education is a junior-college degree. Managers stay in their jobs for a substantial period of time, with median experience at the company being 7 years. Managers do switch gas stations periodically, with median tenure at a gas station of about 2 years. The median number of

[5]Of the remaining five regions, some only record data on a quarterly basis, and others were not made available due to idiosyncratic bureaucratic factors.

employees is 5, so managers have some people management duties, but not for very large groups of workers. The median number of competitors within the local market is 2. The company's definition of the local market starts with a 2.5 km radius around the station as a guideline, but then calls for adjusting the assessed number and type of competitors based on considerations like commuting routes and other factors. The median market share for a station from our partner company, in terms of fuel sales in the local market, is about 30 percent.

Table 2: Descriptive Statistics: Managers and Stations

Manager descriptives		Station descriptives	
Median age	39	Median number of employees	5
Female	34%	Median number of competitors	2
Education level:		Median market share	30%
High school	26%		
Junior college	45%		
College or above	28%		
Median experience (years)	7		
Median tenure at current station (years)	2		

Notes: This table reports descriptive statistics based on the 2024 survey wave. "Experience (years)" represents the total number of years as a station manager, while "tenure at current station (years)" represents the number of years as the station manager at their current assignment. "Competitors" refers to stations from other companies within the local market as defined by the company. "Market share" is a station's share of total fuel sales in its local market as reported by the station manager.

Our survey of district managers shows that station managers have substantial influence over fuel prices, although the degree of autonomy varies across districts. In 48% of the districts, the senior manager reports that station managers can directly change listed fuel prices without needing to submit a proposal (either within a pre-specified range or without restriction). In the remaining districts, managers must make proposals to change listed prices. The survey of district-level managers indicates an average approval rate of about 34 percent. Thus, even when proposals are required, managers can still influence pricing. Overall, station managers have a clear role in influencing listed fuel prices, opening up the possibility that bounded rationality may matter for pricing and performance.

Station managers receive a base salary and substantial performance-based pay, which accounts for roughly 50% of total compensation. Bonuses are tied to two key performance indicators (KPIs): fuel profits and non-fuel (convenience store). Performance is assessed relative to targets, with weighted contributions to bonus pay. These incentives give managers an incentive to choose fuel prices to optimize station profits.

3.1.3 The relevance of intertemporal substitution in fuel markets

It makes sense that intertemporal substitution should be relevant in fuel markets. Consumers face a fixed cost of going to a station to get fuel, and thus have an incentive to

go when their tank is close to empty and then fill up the tank. But they can presumably shift the timing of when they go to some degree, to take advantage of discounts.

As a way to verify whether intertemporal substitution is relevant, we used a fourth dataset. This consists of all transactions by loyalty card holders at all stations of our partner in one region. Transactions are recorded on a minute-by-minute basis. The time period is 2013 to 2015, which unfortunately precludes meaningful overlap with our manager survey datasets. But the data allow a simple approach to testing for the existence of intertemporal substitution: We look at stations who are observed to introduce a policy of a regular weekly discount on listed fuel prices, and see how this affects timing of purchases by day of the week. We find that roughly 13 percent of stations try this policy at some point, usually implementing the discount on Saturday. We identify customers who have purchased at least twice per month in recent months before the policy as regulars, and then compare the timing of their purchases before and after the introduction of the discount policy.

Appendix Figure C.1 shows clear evidence of intertemporal substitution at stations that offer the discount on a single day of the week. Before the policy is introduced, regular customers buy 13.7 percent of their weekly fuel volume on the day that later becomes the discount day, close to the no-shift benchmark of one day in seven. During the policy, this share rises to 29.0 percent. The increase of about 15 percentage points means that this share roughly doubles. The share of weekly transactions on the discount day shows a very similar increase, indicating that regulars mainly buy more often on the discount day rather than buying more per purchase. The increase in sales and number of transactions extend to all treated stations, which include stations with two discount days per week. Among all treated stations, the share of weekly fuel volume per discount day rises by 13.4 percentage points and the share of transactions by 11.8 points. If a manager attributes the higher number of customers on the discount day(s) entirely to attracting new customers, this could imply a large overestimate of the elasticity of demand. These findings are the tip of the iceberg, and intertemporal substitution can occur with other types of price cuts as well, e.g., more prolonged discounts, one-time discounts that are not regularly occurring, etc. In the case of unexpected price cuts there is of course less scope for sales to drop before the discount, but as with regularly occurring discounts, customers can decide to buy earlier than they would have, causing an initial spike in sales but then a temporary dip below normal sales.

3.2 Measuring managers' mental model of consumer response to price discounts

To measure whether a gas station manager has a mental model of consumer response to price discounts that includes IS, we designed a question that presents a hypothetical scenario in a different industry – online laundry detergent sales. The question presents managers with the following scenario and data:

Managers were told that the cost is 4.3 per kg in local currency, and that Retailer A's current pricing strategy involves setting a regular price of 5.9 per kg with a large discount to 5.2 per kg on one randomly selected day each week (days 5 and 11 in the data). The instructions explicitly note that Competitor B does not respond, and the data also show competitor price remaining constant. The column on daily sales shows a clear pattern: sales fall below the baseline of 100 (10,000 kg) before the discount days, but then spike to 202 on discount days, but again fall to 82-91 in the days immediately following the discount, before returning to the baseline of 100. Managers were then asked:

"If retailer A always charges 5.2 per kg, what will the daily sales volume be?"

They were given four response options: 192 (10,000 kg), 165, 100, 202.

The design of the four response options allows us to classify managers' mental models:

- **202**: Managers who choose this option neglect intertemporal substitution. They see the high sales on discount days and assume this represents the true demand at the lower price, without understanding that much of this volume comes from customers shifting their purchase timing.
- **165**: This is the correct answer. Managers who choose it understand that when the discount becomes permanent, customers no longer have an incentive to shift purchase timing, so sales will be lower than on temporary discount days.
- **192**: This answer underestimates intertemporal substitution. These managers understand that some customers shift purchase timing but underestimate how much. More precisely, it corresponds to subtracting the pre-discount dip but not the post-discount dip.
- **100**: These managers attribute the entire demand response to IS. They ignore the true increase in demand due to low prices.

The main reason why we chose this multiple choice format, rather than an open-response format, was to lower cognitive costs of managers in calculating the correct answer. The correct answer, 165, requires subtracting the shifted demand from the discount-day spike. Offering this as one of the response options was intended to help managers who had the correct model arrive at the correct answer. A manager who neglects IS, by contrast, should not understand why this number is one of the options, and choose 202. A potential concern, of course, is that respondents might be inferring explanations for the data from the menu of answer choices, rather than relying on their own mental model of consumer behavior. For this reason, one of the robustness checks in our online experiments involves testing an open response format. All of our main findings go through with this alternative elicitation approach.

A crucial feature of our detergent task design is that it presents managers with a fully controlled scenario where we, as researchers, know the true data-generating process. As we explicitly tell managers in the survey: "This data set is created by us (questionnaire designers) according to specific rules. These rules are designed to simulate the core elements in the real historical data of the industry in a simplified way. You need to infer these rules by observing the data and correctly answer which pricing strategy can bring the best profit under these rules." We further emphasize that "the data shown below is representative and covers all possible situations" and "the table shows all the variables that may affect the profit." This design choice is deliberate and important: unlike real market situations where managers might possess local knowledge or understand market-specific factors that researchers cannot observe, the detergent task contains no hidden information. All relevant factors are visible in the data presented. This allows us to rule out alternative explanations for incorrect answers, such as managers knowing about competitor strategies, local demand shocks, or market-specific dynamics that we as researchers cannot observe. When a manager chooses 202 instead of the correct answer of 165, we can confidently attribute this to neglect of intertemporal substitution rather than to superior local knowledge or consideration of unobserved factors. The controlled nature of this task thus provides clean identification of managers' mental models regarding intertemporal substitution in demand.

3.3 Results from the field: Persistence of wrong mental models, and mechanisms

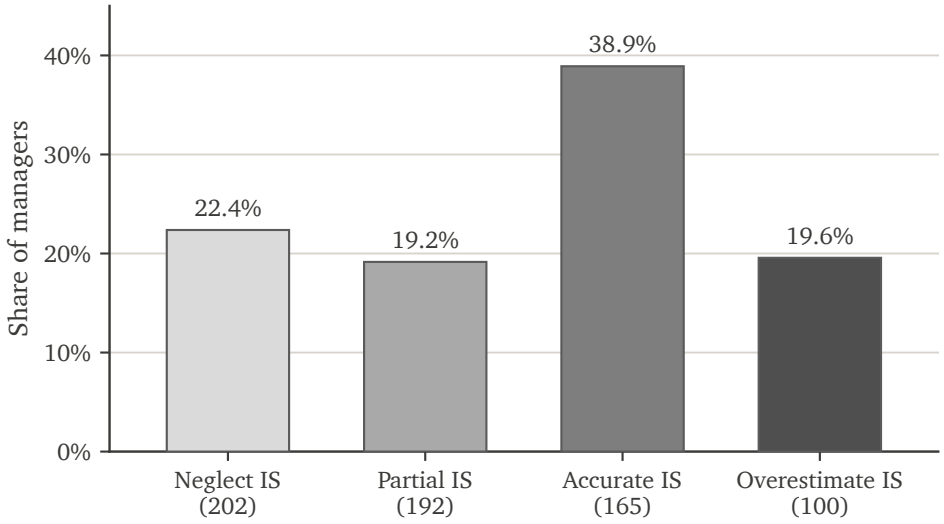
3.3.1 Prevalence of neglect of intertemporal substitution among station managers

We begin by examining whether managers have mental models that neglect intertemporal substitution, using the detergent task. Figure 1 shows the distribution of responses to the question asking managers to predict daily sales if the retailer permanently charges the discounted price.

The results reveal substantial heterogeneity in managers' mental models of consumer response to price. More than 22% completely neglect IS, predicting that sales would remain at the peak level of 202 seen during temporary discounts. Another 19.2% predict 192, consistent with underestimating intertemporal substitution. Roughly 38.9% of managers correctly identify that permanent discounts would result in daily sales of 165. 19.6% attribute the entire sales response to IS, ignoring that there can be a true increase in demand.

If managers neglect IS in their mental models of consumer behavior, we would also expect them to view price cuts as very effective at attracting new customers and thus increasing profits. We checked whether the prediction in the detergent task is, in fact, related to thinking that the detergent retailer could increase overall profits by charging the low price of 5.2 more often. We asked the managers: "What pricing policy do you think would lead to the

Figure 1: Distribution of responses to the detergent task



Notes: This figure shows the distribution of managers' answers to the detergent task, which asks managers to predict daily sales if the retailer permanently charges the discounted price. An answer of 202 neglects intertemporal substitution (IS), 192 underestimates intertemporal substitution, 165 is the correct answer, and 100 attributes the entire sales response to intertemporal substitution. The sample includes $N = 12{,}072$ managers.

highest total profits over the 14 day periods for Retailer A?" Managers could choose from the following options: "The retailer's current pricing policy;"A policy of always charging 5.9;" A policy of always charging 5.2;" Charging the price of 5.2 more frequently than the current pricing policy, but not always 5.2." We define a choice of the first option as favoring a higher pricing strategy, and the third or fourth option as favoring a lower pricing strategy.

Results are consistent with the hypothesized link, and our interpretation of the detergent task. As shown in Appendix Figure D.2, we find that favoring lower prices for the detergent seller is monotonically decreasing across the predictions 202, 192, 165 and 100 in the detergent task, while favoring higher prices is monotonically increasing across them. We provide more details in Appendix D.

Combining managers who predict 202 (22.4%) with those who predict 192 (19.2%), we find that 41.5% of managers underestimate intertemporal substitution in customer behavior. This substantial proportion suggests that neglect or underestimation of intertemporal substitution is widespread among fuel station managers, despite their extensive experience in fuel markets, where discounts are frequent. Correlating an indicator for neglecting IS with managerial experience in years, we find a statistically significant but economically small correlation. Appendix Figure D.3 shows the share of managers who predict 202 by years of experience as a station manager: it is 24 percent among managers with up to two years of experience and 21 percent among managers with 16 or more years.

3.3.2 Evidence on model-directed attention as a mechanism for manager neglect of intertemporal substitution

What types of mental models do managers have if they neglect IS, and what can explain why managers neglect IS, even when shown data involving a stark pre-discount dip and post-discount dip? To shed light on these questions, we asked managers for their explanation of the detergent data in an open-ended question. The question came directly after they saw the data table, and asked:

> "Looking at days 4-7 and days 10-13, why do you think sales show this pattern before and after the days when price is 5.2? Please answer in complete sentences."

Responses can shed light on the role of attentional mechanisms in the sense that mentioning a pattern in the data is a sign of noticing this pattern. A caveat is that the measure likely provides a lower bound on noticing, since managers may not bother to mention some patterns they noticed. For similar reasons, open-ended responses are also likely a lower bound on the prevalence of not neglecting intertemporal substitution, since some managers who do not neglect it may take the path of least resistance of giving a simpler answer.

To classify responses we first looked at a random sample of 1,000 responses, and developed a rubric based on the different categories of explanations found in the sample. The rubric included a category for mentioning IS, in the form of customers changing the timing of their purchases to take advantage of the discount. Other categories included explanations that simply state that the data show how sales increase when prices are high, and more complicated wrong explanations, such as saying that consumer demand is cyclical and the retailer uses discounts at the troughs of the cycle to stimulate demand. The rubric also involved classifying, for each explanation, whether it mentioned the pre-discount dip, the discount-day spike, and the post-discount dip. We then gave the rubric to a large language model in the form of a prompt, and asked the model to classify each of the 13,000 open text responses, both in terms of category of explanations and types of patterns mentioned.

Answers differ in which patterns in the data they mention, as the following two answers show (our translations). A manager who predicted 202 wrote:

> "The price reduction stimulated customers' purchasing desire, leading to increased sales."

A manager who predicted 165 wrote:

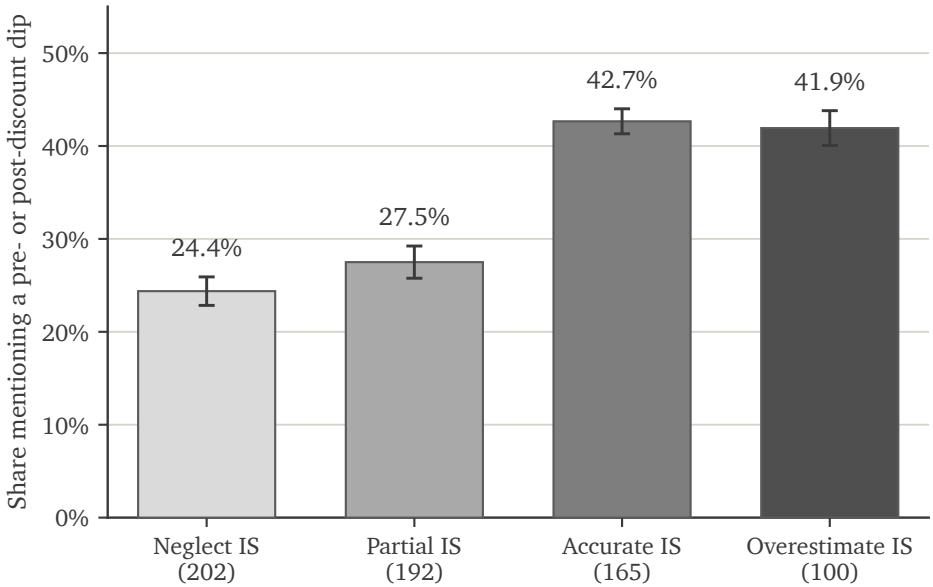
> "Customer traffic decreased the day before and two days after the promotion, as purchases were concentrated on the discount day."

The first answer mentions only the high sales at the low price. The second mentions the pre-discount dip and the post-discount dip and explains them by customers concentrating their purchases on the discount day.

The most frequent type of explanation, provided by 28.5 percent of managers, simply comments on how low prices lead to high sales. By contrast, only 17.2 percent of managers explicitly mention some component of IS. Predicting 165 or 100 in the detergent task is significantly related to mentioning IS in the explanation (Wilcoxon rank-sum test; $p < 0.001$). Having an explanation mentioning IS is highly diagnostic of predicting 165 or 100 in the detergent task: 75% of those who mention IS predict 165 or 100. Conversely, however, only about 22 percent of managers who predict 165 or 100 explicitly mention IS in their explanation. One explanation is that managers who do not underestimate intertemporal substitution may not be able or willing to articulate this in an open-ended question.

Figure 2 shows, for each prediction in the detergent task, the share of managers whose explanation mentions the pre-discount dip or the post-discount dip. Managers who predict 202 are less likely to mention a dip: 24.4 percent of them do, compared to 42.7 percent of managers who predict 165 and 38.8 percent of managers who make any other prediction (both $p < 0.001$). This difference is consistent with model-directed attention: managers whose prediction reveals the wrong model are less likely to mention the patterns the wrong model deems irrelevant.

Figure 2: Mentions of the dips in manager explanations, by prediction



Notes: This figure shows the share of managers whose explanation of the detergent data mentions the pre-discount dip or the post-discount dip, by prediction in the detergent task. Error bars show 95% confidence intervals. The p-values in the text are from two-sample tests of the difference in shares.

Turning to further evidence on mechanisms through which misspecified models may be persistent, Appendix Figure E.4 shows that the most common pattern mentioned is the discount-day spike. 70 percent of these are managers whose explanation is the simple observation that low prices increase sales. This is consistent with the wrong model directing these managers' attention to the discount days, so that they do not notice the pre-discount dip

and the post-discount dip. This could help explain why managers have not learned about intertemporal substitution from seeing data.

In the next section we continue our investigation of mechanisms that can lead to the formation, and persistence, of misspecified mental models of consumer behavior in a set of online experiments.

4 Evidence from online experiments

The previous section shows that the most common explanation of the detergent data among managers who underestimate intertemporal substitution mentions only the discount-day spike. To test whether their ignorance of the dips around the discount days is caused by their mental models or is merely correlated with them, we turn to online experiments, where we randomly assign different models to subjects. In the experiments, subjects on Prolific see the detergent data shown to the managers (Table 1) after reading a "model", a short description of the relationship between sales and prices written by a subject in our initial pilot. Study 1 randomizes which model they read, a description of the wrong model or a description of the correct model. It tests whether the wrong model directs attention away from the dips (Assumption 1), whether subjects who read it predict 202 with the data in view (Prediction 1), and whether directing attention to the dips dislodges it (Prediction 3). Study 2 gives every subject both descriptions, one after the other, and tests whether the effect of the second description depends on which description came first (Prediction 4). Both studies were pre-registered on AsPredicted (#300186 and #300209).

4.1 Study 1

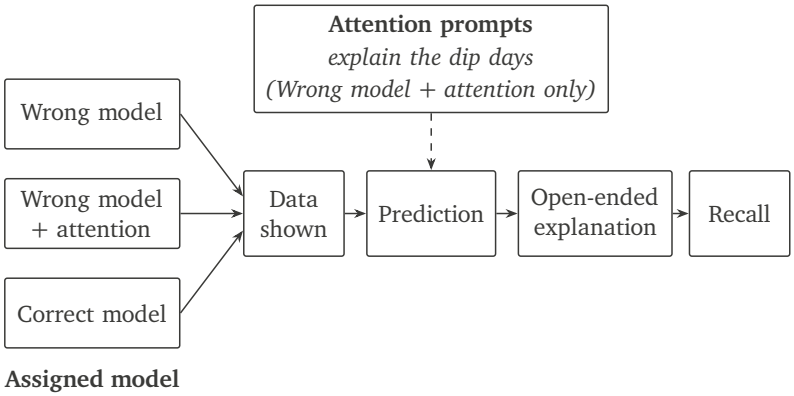
4.1.1 Design

Study 1 has three treatments: *Wrong model*, *Correct model* and *Wrong model + attention*. The treatments differ in the assigned model and, in *Wrong model + attention*, in two attention prompts that subjects answer before the prediction. Otherwise all subjects see the same data and answer the same questions in the same order. Figure 3 shows the timeline.

The instructions describe the data as the manager survey does. Subjects learn that the data come from a market for laundry detergent, that the study designers created the data from a set of rules that mimic the forces in the real industry, that the data are representative and include every variable that affects Retailer A's sales, and that Retailer A announces its pricing schedule to customers at the beginning of each week. They are asked to infer the rules from the data.

The treatment is the assigned model. Before the data appear, subjects learn that others in an earlier part of the study each saw data from many more weeks and wrote down a

Figure 3: Timeline of Study 1



Notes: This figure shows the timeline of Study 1. Subjects are randomized into three treatments: *Wrong model*, *Wrong model + attention* and *Correct model*. The treatments differ in the randomly assigned model and in whether attention prompts are presented; the dashed arrow marks the step that only *Wrong model + attention* receives. All three treatments see the same data and answer the same questions in the same order.

description of the rules relating Retailer A's price to its sales. These descriptions are kept in a bag, and each subject draws one without looking. The page adds that having seen more data gives the writer an advantage in understanding patterns, but does not guarantee that the description is correct. Subjects in *Wrong model* and *Wrong model + attention* draw a description of the wrong model:

> "When the price is low, sales are greater than when the price is high. This has to do with the rule of supply and demand. When the prices are low demand increases."

Subjects in *Correct model* draw a description of the correct model:

> "When retailer A offers a low price on just one day of the week, sales surge on that discount day but drop on the surrounding days, while the other days stay at their usual level. This happens because customers postpone purchases to wait for the discount and then have already stocked up afterward, briefly reducing demand."

Both descriptions are explanations written by subjects in our initial pilot, who saw a series of weekly datasets with different pricing schedules and described the rule relating price to sales. The description is repeated above the data table, so subjects can look it up while they answer.

In *Wrong model + attention*, two attention prompts appear below the data table before the prediction:

> "What do you think explains the pattern in Retailer A's sales from Day 3 to Day 4, and from Day 9 to Day 10?"

"What do you think explains the pattern in Retailer A's sales from Day 6 to Day 8, and from Day 12 to Day 14?"

The first prompt points at the pre-discount dip and the second at the post-discount dip. Subjects are asked to answer each prompt in complete sentences. The prompts tell subjects which days to explain, but they do not say what happens on those days or suggest a reason. They only serve to direct subjects' attention to the dip days.

With both the assigned model and the data in view, subjects answer the prediction question from the manager survey: "Suppose Retailer A set its price to 5.2 every single day (a permanent discount), with Retailer B's price unchanged. What would Retailer A's daily sales be?" They choose among the four answers of the manager survey and earn $1.00 if they choose the correct answer, 165. Subjects then distribute 100 points across the four answers according to how likely they think each one is. This allocation is not incentivized, and we use the points placed on 202 as a measure of confidence in the wrong model. On the next page, without the data, subjects are asked to describe, in complete sentences, the relationship between Retailer A's price and its sales and the mechanism that produces it. One of the two answers is drawn and earns $1.00 if it is graded as correct or mostly correct.

We measure attention by recall.[6] Right after the written explanation, and still without the data, subjects fill in Retailer A's sales for Days 1 to 7, or tick "I don't recall" for any day. They start with $0.35, gain $0.10 for each correct value and lose $0.05 for each incorrect value, while "I don't recall" costs nothing. We code whether a subject recalls a pattern rather than an exact number, because model-directed attention concerns whether a day is noticed as different from a baseline day, and this coding is pre-registered. A subject recalls *the discount-day spike* if the value entered for Day 5 is above 100, *the pre-discount dip* if the value entered for Day 4 is below 100, and *the post-discount dip* if the value entered for Day 6 or Day 7 is below 100. As a placebo, we code whether a subject enters a value below 100 for any of Days 1 to 3, when sales were 100.

The study ends with a 9-item Raven progressive test as the measure of cognitive skills, and a demographic survey. While the full version of Raven progressive matrices includes 60 questions, the abbreviated 9-item version has been shown to be a reliable proxy for cognitive skills (Bilker et al., 2012). The test is also incentivized: subjects receive $0.5 for each correct

[6]We assume that the assigned model affects what subjects encode, and not how they store or retrieve what they have encoded. The assumption is consistent with Gagnon-Bartsch, Rabin and Schwartzstein (2026), where the model directs attention at the moment the data are seen. Recall is elicited after the data are removed, so decay pushes recall of every pattern down and works against finding a difference across treatments. Thus treatment effect on recall is potentially conservative measure of the treatment effect on attention. While measurement error is a concern with the recall measure, we also complement it with two other measures. In an earlier pilot we also measured attention through the "top-down" approach, by hiding each sales value behind a button that revealed it for one second when clicked, so that clicks on a day count as attention to that day. Appendix G.1 reports the results: subjects assigned the wrong model click the dip days less than subjects assigned the correct model. Further, the causal effect of attention is measured by the attention prompts, which change attention without relying on either measure of it.

answer. The demographic survey includes basic demographic information, such as gender, age and education.

We recruited US adults on Prolific in July 2026. Of the 600 subjects who completed the study, we exclude those who ticked "I don't recall" for all seven days in the recall task, which we see as a sign of low effort, and those whose answers were flagged as generated by an AI tool. Both exclusions are pre-registered. They leave 549 subjects: 177 in *Wrong model*, 180 in *Wrong model + attention* and 192 in *Correct model*. The median subject took 15 minutes.

4.1.2 The assigned model and attention to the data

We first test whether the assigned model causally directs subjects' attention to different data patterns as predicted by the conceptual framework. In particular, we examine whether the wrong model directs attention away from the pre-discount dip and the post-discount dip compared to the correct model, while not lowering attention to the discount-day spike.

Table 3 shows how the typical recalled week of sales differ between different assigned models. In particular, it reports, for each of the first seven days, the median value recalled by subjects, excluding entries of "I don't recall" and those below 20 (the actual minimum sales is 82) and above 400 (the actual maximum is 202). Subjects in *Correct model* recall a week close to the data, with medians of 92, 90 and 93.5 on Days 4, 6 and 7. In *Wrong model*, the median is 202 on Day 5 and 100 on every other day. On Days 6 and 7 most answers in *Wrong model* are exactly 100, while on Day 4 the answers split almost evenly between 100 and values below it.

Table 3: The typical recalled week in Study 1

	Data shown		Median recalled sales	
Day	Price	Sales	*Wrong model*	*Correct model*
1	5.9	100	100	100
2	5.9	100	100	100
3	5.9	100	100	100
4	5.9	**90**	100	92
5	5.2	202	202	202
6	5.9	**82**	100	90
7	5.9	**91**	100	93.5

Notes: This table reports, for each of Days 1 to 7, the price and sales shown to subjects and the median value entered in the recall task by subjects who did not tick "I don't recall" for that day. Entries below 20 or above 400, mostly prices typed into the sales boxes, are excluded. Shaded cells mark the three dip days, Days 4, 6 and 7. The sample includes $N = 177$ subjects in *Wrong model* and $N = 192$ in *Correct model*.

The wrong model lowers recall only of the patterns it deems irrelevant. Figure 4 shows the share of subjects who recall each pattern. Recall of the pre-discount dip is 27 percent in *Wrong model*, compared to 41 percent in *Correct model*, and recall of the post-discount

dip is 27 compared to 55 percent (both $p < 0.01$). Almost everyone recalls the discount-day spike, 94 percent in *Wrong model* compared to 88 percent in *Correct model* ($p < 0.10$), and the placebo does not change: 18 percent in *Wrong model* and 20 percent in *Correct model* enter a value below 100 for one of the first three days. Subjects in *Wrong model* tick "I don't recall" more often on all six regular-price days, by 13 percentage points on Days 1 to 3. The additional difference on the dip days is a swap. Relative to Days 1 to 3, the share of subjects in *Wrong model* who enter a value below 100 on a dip day is 19 percentage points lower and the share who enter exactly 100 is 19 percentage points higher (both $p < 0.01$), with no additional difference in "I don't recall."

One concern with the high rate of dips recalled in *Correct model* is that the correct model itself mentions the two dips, and subjects recalled the description of the model, not the data. A more detailed look at the exact recalled values suggests otherwise. Subjects in *Correct model* enter the exact value of at least one dip day more often than subjects in *Wrong model*, 29 compared to 19 percent ($p < 0.05$). At the level of individual answers, 51 percent of subjects in *Wrong model* enter exactly 202 for Day 5 and no value below 100 for any other day, compared to 20 percent in *Correct model* ($p < 0.01$). Half of the subjects who read the wrong description recall the discount-day spike and nothing that separates one regular-price day from another, consistent with the noticing strategy of Prediction 1.

Result 1. *The wrong model lowers recall of the pre-discount and the post-discount dips, which it deems irrelevant, but does not lower recall of the discount-day spike, which it explains.*

4.1.3 The assigned model and predictions

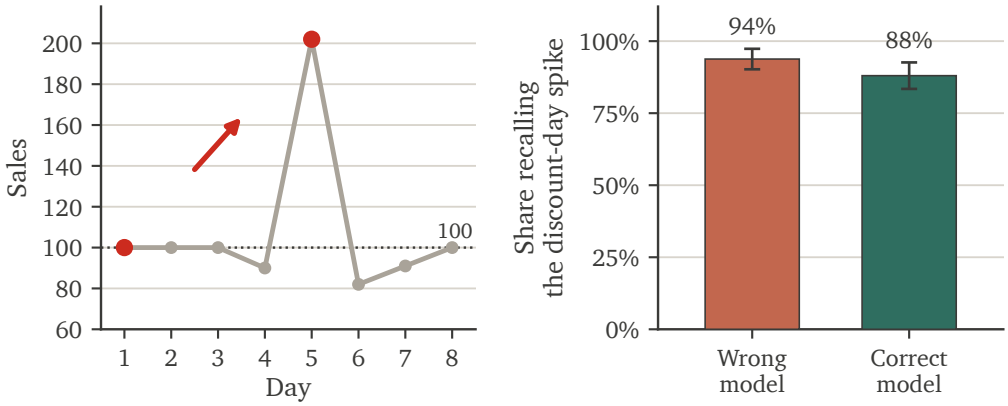
The wrong model also changes what subjects predict. Figure 5 shows that 74 percent of subjects in *Wrong model* predict 202, compared to 32 percent in *Correct model*, a difference of 42 percentage points ($p < 0.01$). Both treatments see a table in which sales fall below 100 around each discount. Subjects in *Wrong model* place 47 of their 100 points on 202 on average, compared to 23 in *Correct model* ($p < 0.01$). In *Correct model*, 41 percent choose 165 and 20 percent choose 100; in *Wrong model*, 15 percent choose 165.

A menu of four answers could suggest a model to subjects who would not otherwise have one. In a follow-up pilot of Study 1 with 229 subjects, subjects type their prediction into an empty box (Appendix G). The wrong model again raises the share typing exactly 202, from 23 to 48 percent, and the share typing 200 or more, from 33 to 67 percent (both $p < 0.01$). It again lowers recall of the two dips and leaves recall of the discount-day spike and the placebo unchanged.

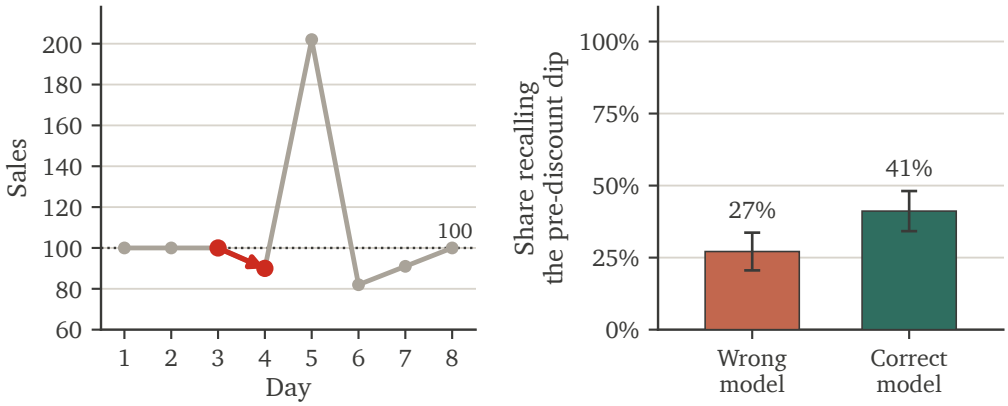
Result 2. *The wrong model raises the share of subjects who neglect intertemporal substitution.*

Figure 4: Recall of the three patterns in Study 1: *Wrong model* and *Correct model*

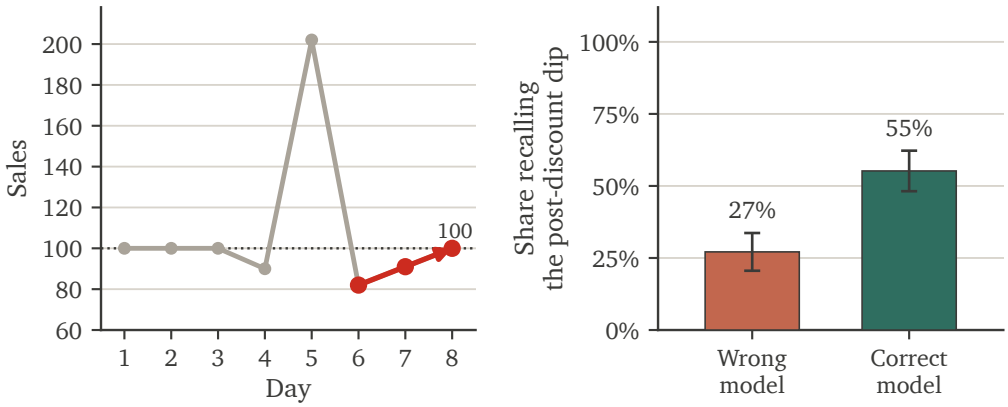
(a) The discount-day spike



(b) The pre-discount dip

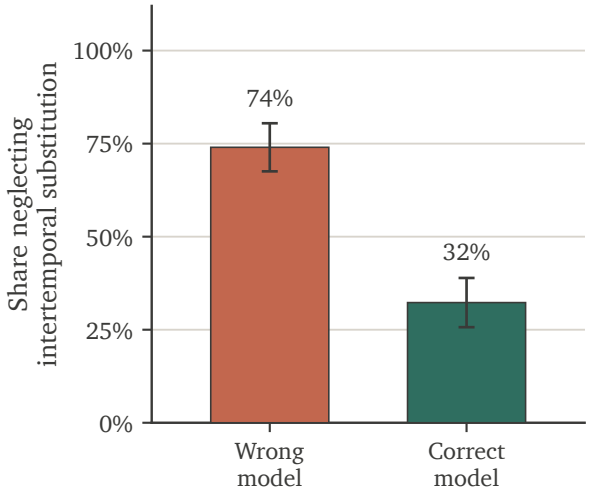


(c) The post-discount dip



Notes: This figure shows how often subjects recall each of the three patterns. Panel (a) shows the discount-day spike, Panel (b) the pre-discount dip and Panel (c) the post-discount dip. In each panel, the left plot shows sales on Days 1 to 8 and marks one pattern with an arrow, and the right plot shows the share of subjects who recall that pattern. A subject recalls the discount-day spike if the value entered for Day 5 is above 100, the pre-discount dip if the value entered for Day 4 is below 100, and the post-discount dip if the value entered for Day 6 or Day 7 is below 100. Ticking "I don't recall" counts as not recalling the pattern. The sample includes $N = 177$ subjects in *Wrong model* and $N = 192$ in *Correct model*. Error bars show 95% confidence intervals.

Figure 5: Share neglecting intertemporal substitution in Study 1: *Wrong model* and *Correct model*



Notes: This figure shows the share of subjects whose best guess of daily sales under a permanent price of 5.2 is 202, the answer that neglects intertemporal substitution. The correct answer is 165. The sample includes $N = 177$ subjects in *Wrong model* and $N = 192$ in *Correct model*. Error bars show 95% confidence intervals.

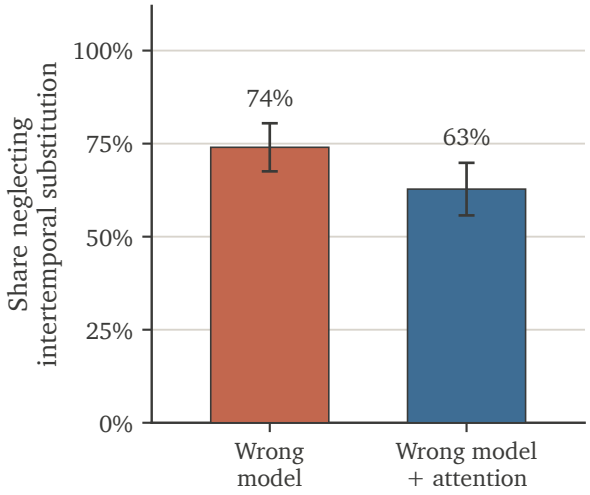
4.1.4 Directing attention to the dips

The attention prompts in *Wrong model + attention* restore recall of the dips. Appendix Figure G.10 compares this treatment to *Wrong model*. Recall of the pre-discount dip rises from 27 to 49 percent and recall of the post-discount dip from 27 to 59 percent (both $p < 0.01$). Both shares are close to those in *Correct model*, 41 and 55 percent, and neither difference from *Correct model* is significant. The prompts do not raise the placebo, 23 compared to 18 percent. Recall of the discount-day spike moves in the opposite direction to recall of the dips: it is 94 percent in *Wrong model*, 88 percent in *Correct model* and 84 percent in *Wrong model + attention*. The assigned model and the attention prompts change which days subjects recall.

This treatment also addresses the possibility that subjects fill in the week from their model at the time of recall. Both treatments read the wrong description, and among subjects who predict 202, 67 percent in *Wrong model + attention* recall at least one dip, compared to 37 percent in *Wrong model* ($p < 0.01$). Subjects who give the wrong model's answer recall the dips when the prompts directed their attention to those days.

The attention prompts also lower the share of subjects who predict 202. Figure 6 shows that the share falls from 74 percent in *Wrong model* to 63 percent in *Wrong model + attention*, a reduction of 11 percentage points ($p < 0.05$). Subjects in *Wrong model + attention* also place fewer points on 202, 39 compared to 47 ($p < 0.05$), and with controls for age, gender and Raven score the reduction in the share predicting 202 is 13 percentage points ($p < 0.01$). Asking about these days tells subjects that they are worth explaining, but the prompts state no pattern and no explanation. Because the assigned model is the same in both treatments, we interpret the reduction as the effect of noticing the dips.

Figure 6: Share neglecting intertemporal substitution in Study 1: *Wrong model* and *Wrong model + attention*



Notes: This figure shows the share of subjects whose best guess of daily sales under a permanent price of 5.2 is 202, the answer that neglects intertemporal substitution. The correct answer is 165. The sample includes $N = 177$ subjects in *Wrong model* and $N = 180$ in *Wrong model + attention*. Error bars show 95% confidence intervals.

The prompts bring recall of the dips to the level of *Correct model*, but not the prediction. In *Wrong model + attention*, 63 percent of subjects predict 202, compared to 32 percent in *Correct model* ($p < 0.01$). The subjects who move off the wrong model do not move to 165: the share choosing 165 is 20 percent, compared to 15 percent in *Wrong model*, a difference that is not significant, while the share choosing 100 rises from 2 to 7 percent ($p < 0.05$). Section 2 offers a potential explanation for why noticing the dips need not change the answer. A subject who attributes the dips to the calendar, with some days simply selling less, holds the predictor-neglect model, which contains no intertemporal substitution and still gives 202 as the answer.

Result 3. *Directing attention to the dips lowers the share of subjects who neglect intertemporal substitution.*

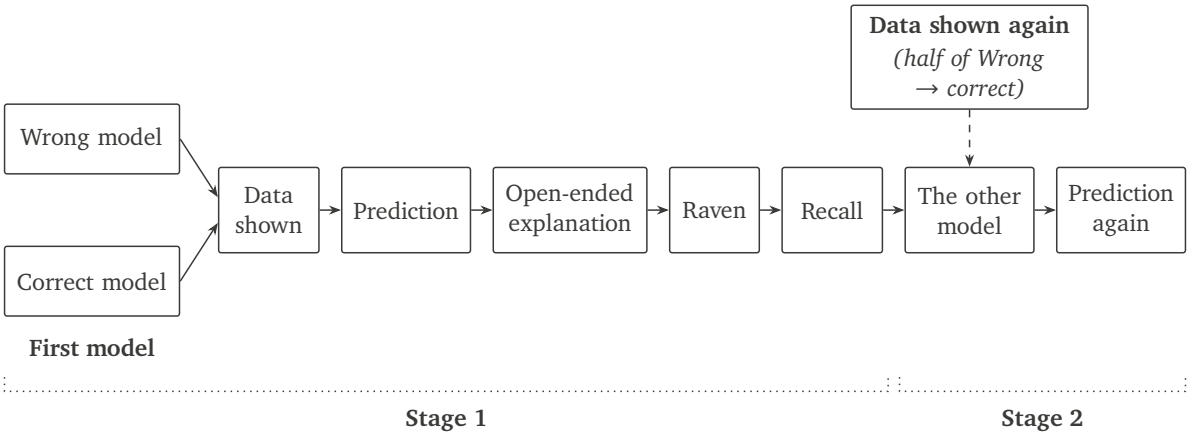
4.2 Study 2

4.2.1 Design

Study 2 asks what happens when subjects who have already formed a view of the data encounter the other model. The study has two stages, and Figure 7 shows the timeline. From the first page, subjects know that they will draw one description from the bag now and a second description from the same bag later. The second description therefore cannot be mistaken for a response to their own answer.

Stage 1 repeats Study 1 without the attention prompts. Subjects draw a first description, see the data, predict sales under a permanent price of 5.2, distribute 100 points across

Figure 7: Timeline of Study 2



Notes: This figure shows the timeline of Study 2. Subjects are randomized into three treatments in the ratio 2:2:1. The two lanes differ in which model is assigned first; the dashed box marks the third treatment, in which half of the subjects assigned the wrong model first see the data again in Stage 2.

the four answers and write the explanation. They then answer the nine Raven matrices and, after that, recall the first week of sales, with the same incentives as in Study 1. The Raven block places time and an unrelated task between the data and the recall. In stage 2, subjects draw a second description, which is always the other one: subjects assigned the wrong model first now read the correct description, and subjects assigned the correct model first now read the wrong description. Only the second description is shown. Subjects then choose among the same four answers again and distribute 100 points again, without writing a second explanation. They are told that one of the two predictions is drawn for payment.

Subjects are randomized into three treatments in the ratio 2:2:1. In two treatments, subjects are assigned the wrong model first, and in one of these the data table is shown again next to the correct description in stage 2. In the third treatment, subjects are assigned the correct model first and see no data in stage 2. Showing the data again tests whether subjects judge the correct model against their memory of the data. For a subject who recalls the discount-day spike but not the dips, the correct model fits the recalled week no better than the wrong model. If subjects judge the correct model against their memory, showing the data again should lower the share predicting 202, with a larger effect among subjects who recall the discount-day spike but not the dips.

We again recruited US adults on Prolific. We apply the same two pre-registered exclusions as in Study 1 to the 1,003 subjects who completed the study. This leaves 896 subjects: 720 assigned the wrong model first (367 without and 353 with the data shown again) and 176 assigned the correct model first. The median subject took 19 minutes.

4.2.2 Encountering the other model

Stage 1 replicates Study 1. Among subjects assigned the wrong model first, 72 percent predict 202, compared to 28 percent among those assigned the correct model first ($p < 0.01$). Even though recall now comes after the Raven block, subjects assigned the wrong model first recall the discount-day spike far more often than the dips: 86 percent recall the discount-day spike, compared to 28 percent for the pre-discount dip and 30 percent for the post-discount dip ($p < 0.01$). Compared to subjects assigned the correct model first, they recall the post-discount dip less often, 30 compared to 52 percent ($p < 0.01$), and the pre-discount dip less often, 28 compared to 34 percent ($p < 0.10$), while recall of the discount-day spike is 86 percent in both groups.

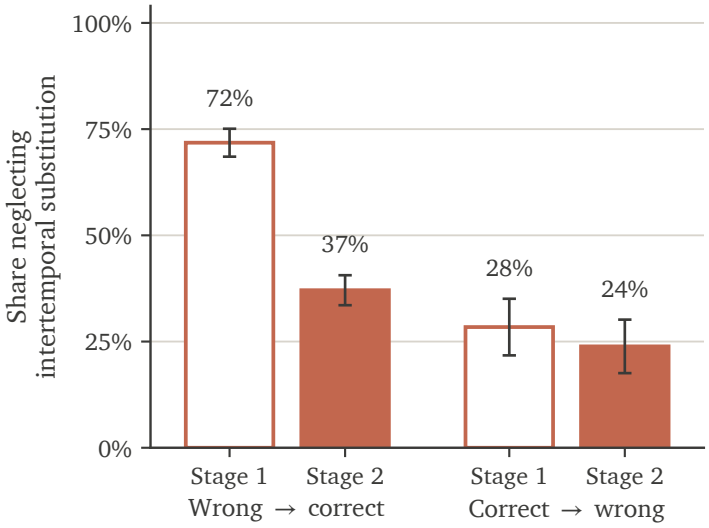
Showing the data again next to the correct description does not change the share predicting 202. Without the data, the share predicting 202 falls by 35 percentage points; with the data, it falls by 34 percentage points, and it is 37 percent at stage 2 in both treatments. The 95 percent confidence interval of the difference between the two changes runs from −6 to 9 percentage points and excludes effects of 10 percentage points in either direction. Showing the data again also has no larger effect among subjects who recall the discount-day spike but not the dips. Because the correct description states in words that sales drop on the dip days, a subject assigned the wrong model first can correct the prediction without returning to the data or to the memory of it. Study 2 therefore does not show that the correction runs through the dips that subjects noticed in stage 1. We pool the two treatments below.

Predictions change in one order only, consistent with Prediction 4. Figure 8 shows the share predicting 202 in each stage. Among subjects assigned the wrong model first, the share falls from 72 to 37 percent after they read the correct description. Among subjects assigned the correct model first, it falls from 28 to 24 percent after they read the wrong description. The difference between the two changes is 30 percentage points ($p < 0.01$). Individual changes in prediction show the same asymmetry. Of the 517 subjects who predict 202 after reading the wrong description, 51 percent move off the wrong model after encountering the correct model. Of the 126 subjects who do not predict 202 after reading the correct description, 4 move to the wrong model after encountering the wrong model. In the framework, subjects who keep predicting 202 after encountering the correct model place no weight on it.

Result 4. *Encountering the correct model after being assigned the wrong model first significantly moves subjects off the wrong model; however, encountering the wrong model after being assigned the correct model first has no effect on moving subjects towards the wrong model.*

Subjects assigned the wrong model first move off it after encountering the correct model, but the correction is not full. At stage 2, both groups have read both descriptions and differ only in the order in which they read them. Appendix Figure G.11 shows the distribution of predictions in each stage. After both descriptions, 37 percent of subjects assigned the wrong

Figure 8: Share neglecting intertemporal substitution in each stage of Study 2



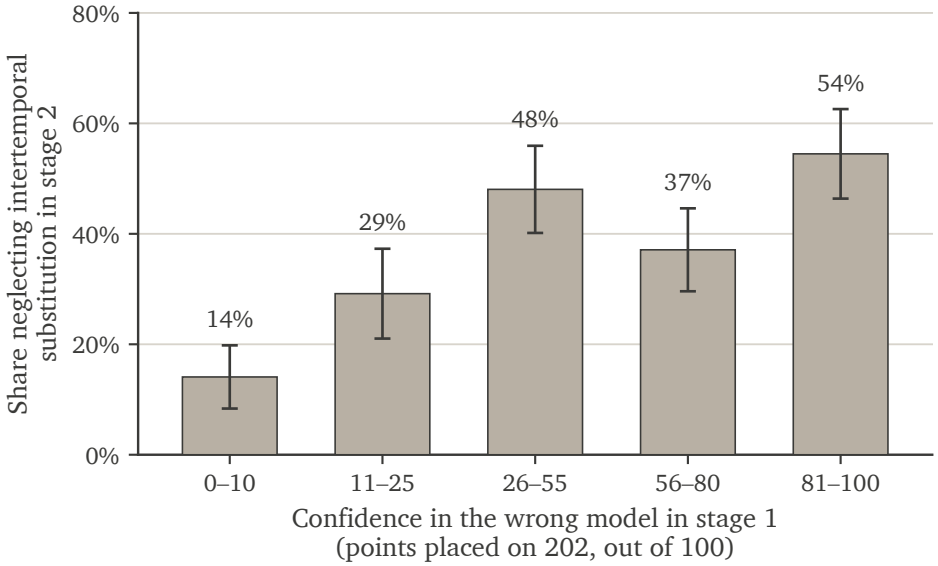
Notes: This figure shows the share of subjects who predict 202, the answer that neglects intertemporal substitution, in stage 1 (hollow bars) and in stage 2 (solid bars). Wrong → correct: subjects assigned the wrong model first and the correct model second, pooling those with and without the data shown again ($N = 720$). Correct → wrong: subjects assigned the correct model first and the wrong model second ($N = 176$). Error bars show 95% confidence intervals.

model first predict 202, compared to 24 percent of those assigned the correct model first ($p < 0.01$), and 35 percent choose 165, compared to 49 percent ($p < 0.01$). The subjects who move off the wrong model do not all move to 165: of the 265 who move off it, 51 percent choose 165, 29 percent choose 192 and 20 percent choose 100. If subjects simply adopted the more detailed description, the correct description would move more subjects than the wrong one, but the two groups would end in the same place once both had read both descriptions. The persistence of the first model indicates that being assigned the wrong model first has a lasting effect even after encountering the correct model.

Result 5. *There is an order effect when both models are assigned: being assigned the wrong model first leads to a higher level of neglect of intertemporal substitution than being assigned the correct model first.*

Section 2 proposes confidence in the first model as a proxy for whether a subject entertains the second, since a subject who is certain of the wrong model sees nothing left to learn from the correct one. Figure 9 plots, for the 720 subjects assigned the wrong model first, the share who predict 202 in stage 2 against the points they allocated to 202 in stage 1. The share is 16 percent among subjects who placed 0 points on 202 and 62 percent among those who placed all 100 points on it. Each additional 10 points on 202 is associated with a share that is 3.6 percentage points higher ($p < 0.01$). The more confident subjects were in the wrong model, the more often they neglect intertemporal substitution after encountering the correct model.

Figure 9: Confidence in the wrong model and the prediction in stage 2 of Study 2



Notes: This figure shows the share of subjects who neglect intertemporal substitution (predict 202) in stage 2, by the points they placed on 202 in stage 1, for subjects assigned the wrong model first ($N = 720$). Subjects are divided into five groups of roughly equal size by the points they placed on 202: 0 to 10 points ($N = 142$), 11 to 25 ($N = 120$), 26 to 55 ($N = 154$), 56 to 80 ($N = 159$) and 81 to 100 ($N = 145$). Error bars show 95% confidence intervals.

5 Consequences of neglecting intertemporal substitution among station managers

5.1 Mental models of intertemporal substitution and understanding consumer response to fuel price cuts

We turn next to investigating whether neglecting intertemporal substitution, as captured by our controlled detergent task, translates into neglecting how fuel price cuts lead to intertemporal substitution. To do this, we present managers with real gas station data that show an intertemporal substitution response to a sustained price cut, and ask them to make counterfactual predictions. We aim to test whether managers who predict 165 or 100 in the detergent task predict an intertemporal substitution response to the cut in fuel prices.

5.1.1 The gas station counterfactual task

Managers are shown the following actual data from a gas station:

The data show a clear pattern consistent with intertemporal substitution. When the price drops from 5.83 to 5.53 on March 5, sales spike to 11,273 liters—a substantial increase from the previous day. However, sales then decline sharply over the following days. The same pattern emerged when price drops from 5.53 to 5.33 on March 14, with it firstly jumps to 16,206 but then falls to 7,572 the next day. This pattern suggests that the (surprise) price cut induced some customers to fill up earlier than planned, borrowing from future demand.

Table 4: Real Gas Station Data Presented to Managers

Date	Price	Price ceiling	Fuel sales	Profit
February 28	5.83	5.83	7,800	7,800
March 1	5.83	5.83	4,275	4,275
March 2	5.83	5.83	6,439	6,439
March 3	5.83	5.83	9,776	9,776
March 4	5.83	5.83	9,293	9,293
March 5	5.53	5.83	11,273	7,891
March 6	5.53	5.83	6,431	4,501
March 8	5.53	5.83	4,676	3,273
March 9	5.53	5.83	2,350	1,645
March 10	5.53	5.83	4,020	2,814
March 11	5.53	5.83	5,875	4,112
March 12	5.53	5.83	6,376	4,463
March 14	5.33	5.83	16,206	8,103
March 15	5.33	5.83	7,572	3,786

Managers were then asked a counterfactual question:

> "If the station had reduced the price to 5.53 per liter starting from March 2 instead of March 5, what do you think the sales volume would have been from March 2 to March 6?"

They were asked to predict sales for each of these five days. This task is more complex than the detergent task for several reasons. First, it uses real data with natural day-to-day variation rather than stylized patterns. Second, it involves a specific market context where managers might consider factors beyond IS, such as local demand patterns or competitor responses. Third, the counterfactual requires managers to reason about how changing the timing of a price cut would affect the temporal distribution of sales.

Despite this complexity, the task provides a valuable test of whether managers can notice intertemporal substitution in realistic market data. If managers understand that price cuts cause intertemporal substitution, they should predict that moving the price cut to March 2 would shift the sales spike earlier, with a similar pattern of subsequent decline. Conversely, managers who neglect intertemporal substitution might predict that sales would simply increase to a higher level but follow the same pattern of spiking on the original price cut day (March 5).

5.1.2 Responses to the gas station counterfactual task

To understand the nature of responses to the task, we first performed a principal component analysis (PCA) on the standardized sales predictions for all five days. Table 5 shows that responses can be summarized by two components having eigenvalues greater than 1. Component 1 loads positively and relatively evenly on all five days except for the discount start, capturing predicting high sales on all days except for the first day. Component 2 loads very strongly on the first day, but has lower and declining loadings on subsequent days.

Given that a prediction that reflects intertemporal substitution has high sales on Day 1, and falling and low sales afterwards, managers whose predictions reflect intertemporal substitution should have a high score on Component 2, and a low score on Component 1.

Table 5: PCA Scoring Coefficients for Sales Predictions

Variable	Component 1	Component 2
Day 1 sales (z-score)	0.048	0.949
Day 2 sales (z-score)	0.504	0.179
Day 3 sales (z-score)	0.549	-0.021
Day 4 sales (z-score)	0.525	-0.256
Day 5 sales (z-score)	0.408	0.026

Notes: This table reports the scoring coefficients from a principal component analysis of managers' standardized sales predictions for the counterfactual scenario.

Using the standardized predicted sales, we next performed k-means clustering to identify groups of managers with similar prediction patterns. Based on the (data driven) Calinski-Harabasz criterion, we identify four distinct clusters. We also obtain 4 cluster using alternative approaches such as visual inspection of an elbow plot.

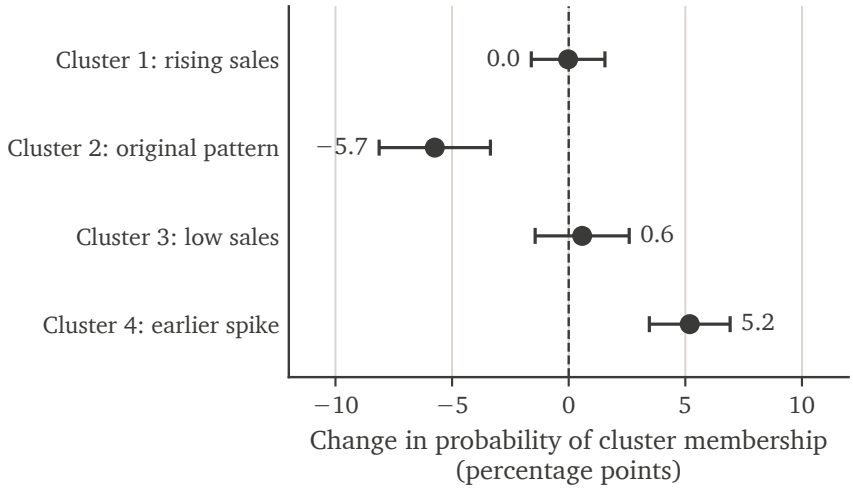
Appendix Figure E.7 visualizes the clusters in the two-dimensional space defined by the principal components, and shows that there is a distinct cluster, Cluster 4, which has high values on Component 2 and low values on Component 1, consistent with a group of managers whose predictions reflect intertemporal substitution. Appendix Figure E.8 plots the average predictions over time for the four clusters, and we see that Cluster 4 does, indeed, predict an intertemporal substitution response: A high value of sales on the new, earlier discount day, and then falling sales afterwards. The heavy line in the figure shows the profile presented in the original data, and reveals that Cluster 4 predicts a level of sales on the new discount day that closely matches the level shown in the data for the original discount day. The other clusters can be summarized as follows: Cluster 3 are managers who on average predict a generally low level of sales over time; Cluster 2 are managers who mainly just reproduce the original pattern with their predictions; Cluster 1 are managers who predict high sales on the new discount day and then high and even increasing sales afterwards.

5.1.3 Relationship of gas station counterfactual predictions to predictions in the detergent task

To investigate the relationship between predicting 165 or 100 in the detergent task and predictions in the gas station counterfactual task, we estimate a multinomial logit. The dependent variable takes on one of four values, corresponding to the clusters identified in our analysis of sales predictions, and the independent variable is an indicator for predicting 165 or 100 in the detergent task.

Figure 10 plots marginal effects from the regression. We see that predicting 165 or 100 is associated with a significantly higher probability of being in Cluster 4 for the gas station counterfactual task, corresponding to predicting an IS response to the fuel price cut. This is consistent with the hypothesized link between predictions in the detergent task and predictions about fuel price cuts. Notably, predicting 165 or 100 is also associated with a significantly lower probability of being in Cluster 2, which is the group of managers who tended to reproduce the original sales profile in the gas station task. It makes sense that predicting 165 or 100 moves a manager from the group who noticed the original pattern in the gas station task, to a group that translates this pattern to the earlier time period because of the earlier discount. Put another way, noticing the original pattern in the gas station task is a pre-condition for predicting a similar IS response to the earlier, counterfactual discount.

Figure 10: Predicting 165 or 100 in the detergent task and sales prediction cluster



Notes: This figure reports average marginal effects, in percentage points, from a multinomial logit of a manager's sales prediction cluster on an indicator for predicting 165 or 100 in the detergent task. A manager is coded 1 if the detergent prediction is 165 or 100, and 0 if it is 192 or 202. Clusters are those of Appendix Figure E.8, and Cluster 2 is the base outcome. The sample includes $N = 6{,}840$ managers, of whom 3,981 are coded 1. The regression includes no controls, and standard errors are robust. Error bars show 95% confidence intervals.

5.2 Mental models of intertemporal substitution and perceived elasticity

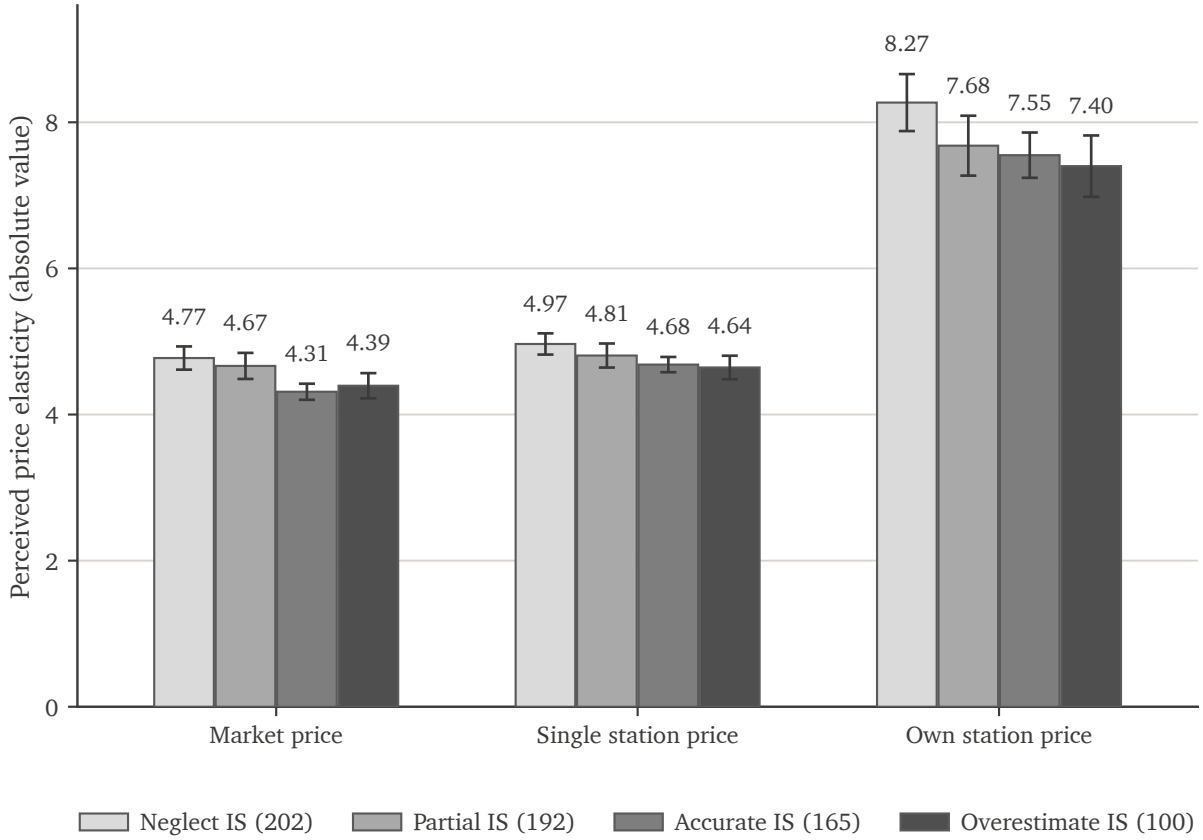
To understand how the prediction in the detergent task affects managers' beliefs about price elasticity of demand, we elicit perceived elasticity through three complementary approaches that capture different pricing contexts. Our first measure asks managers to predict how sales at a "typical station" of our partner company would respond when the price ceiling is reduced by 30 cents. This scenario represents a market-wide price change that affects all stations simultaneously, allowing us to measure beliefs about market-level demand elasticity. Our second measure shifts focus to competitive pricing by asking managers to predict

sales changes when a representative station unilaterally reduces its price by 30 cents while the price ceiling remains unchanged. Managers predict weekly sales for both the first and second week of the price cut. This measure captures beliefs about own-price elasticity, where price cuts can attract customers from rival stations. Our third measure personalizes the scenario by asking managers to make the same predictions for their own station rather than a representative station. This elasticity is arguably more important for their pricing decisions. For all three measures, we calculate implied elasticity using the standard formula: the percentage change in quantity divided by the percentage change in price. In the first two scenarios, we provide the baseline weekly sales of fuel products at this representative station before the price cut.

Figure 11 presents managers' perceived price elasticities under three scenarios on a common axis: a market-wide price change, when the price ceiling falls; a unilateral price cut at a representative single station; and the same unilateral cut at the manager's own station. The results reveal two key patterns. First, managers consistently perceive own-price elasticity to be higher than market price elasticity, reflecting their belief that unilateral price cuts can attract customers from competitors in addition to any market expansion effects. Second, and more importantly, the figure shows a clear monotonic relationship between the prediction in the detergent task and perceived elasticity. For both market and own-price scenarios, managers who predict 202 perceive the higher elasticities, while those who predict 100 perceive lower elasticities. This pattern is particularly pronounced for own-price elasticity, where the difference between the extreme groups is substantial. The monotonic reduction in perceived elasticity across the predictions 202, 192, 165 and 100 confirms that neglecting intertemporal substitution leads managers to systematically perceive demand as more responsive to price changes than it is.

The third scenario in Figure 11 extends the analysis to managers' beliefs about elasticity at their own stations. When asked to predict how their specific station would respond to a 30-cent price cut, managers show two notable patterns. First, comparing these perceived elasticities to the perceived single station elasticities for a representative station, managers consistently believe their own station faces higher elasticity than a typical station. This may reflect overconfidence in their ability to attract customers through price cuts or (biased) beliefs that their particular market is more price-sensitive than average. Second, we again observe the monotonic relationship between the prediction in the detergent task and perceived elasticity. Managers who predict 202 perceive the highest elasticities, followed by those who predict 192, those who predict 165, and finally those who predict 100, who perceive the lowest elasticities. This consistent pattern across different contexts—from abstract detergent markets to representative fuel stations to managers' own stations—demonstrates that neglect of intertemporal substitution fundamentally shapes how managers conceptualize the relationship between prices and demand. Those who neglect intertemporal substitution systematically perceive demand as more responsive to price changes than it truly

Figure 11: Mental models of intertemporal substitution and perceived price elasticities



Notes: This figure shows managers' perceived price elasticities of demand, in absolute value, by their prediction in the detergent task, for three scenarios on a common axis. *Market price* is the predicted demand response when the price ceiling falls by 30 cents. *Single station price* is the predicted response when one station cuts its price by 30 cents while the ceiling is unchanged, for a representative station. *Own station price* is the same unilateral cut at the manager's own station. Error bars show 95% confidence intervals.

is, regardless of whether they are thinking about hypothetical scenarios or their own daily operations.

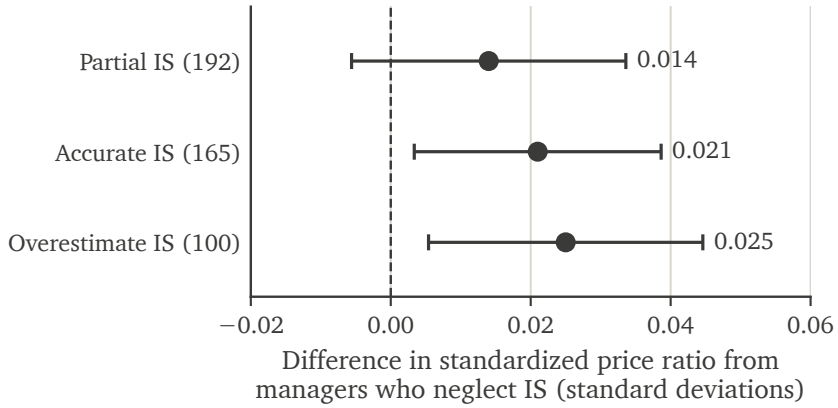
5.3 Mental models and actual pricing behavior

Before examining actual pricing decisions, we first explore managers' beliefs about the profitability of price discounts. Figure E.9 shows responses to a direct question asking managers whether they believe price cuts are good or bad for fuel profits. Managers answer differently in the expected direction depending on whether they underestimate intertemporal substitution (predict 202 or 192) or not (predict 165 or 100). Among managers who underestimate intertemporal substitution, around 16% view price cuts as "always good" for profits, compared to only 12% among those who do not underestimate IS. Conversely, managers who predict 165 or 100 are more skeptical of price cuts: 35% believe they are "mostly bad" for profits, compared to 31% among those who underestimate IS. The shares believing price cuts are "mostly good" and "always bad" are similar between the two groups of managers.

This difference in distribution is consistent with the prediction that managers who underestimate intertemporal substitution are more optimistic about the profit potential of discounts, likely because they overestimate the gain from new customers due to misinterpretation of data.

To examine whether differences in managers' predictions in the detergent task are reflected in actual pricing decisions, we analyze the standardized monthly price ratio (defined in Section 3.1) from January 2019 to October 2024 for approximately 10,000 stations matched to our survey responses, keeping only the months in which the station was run by the manager who answered the survey. We estimate specifications with increasingly comprehensive controls: (1) panel-month-by-district fixed effects only, (2) adding market conditions (competitor counts by type, location, 24-hour operation, rental status), and (3) further incorporating manager characteristics (cognitive skills, gender, age, experience, strategic sophistication).

Figure 12: Mental models of intertemporal substitution and actual prices



Notes: This figure reports coefficient estimates from regressing the standardized monthly price ratio on indicators for the prediction in the detergent task, with managers who neglect intertemporal substitution (predicting 202) as the omitted category. The specification includes panel-month-by-district fixed effects, market conditions (competitors, location type, 24-hour operation, rental status), and manager characteristics (cognitive skills, gender, age, experience, strategic sophistication). The sample is station-months under the manager who answered the survey, January 2019 to October 2024. Standard errors are clustered at the station level. Error bars show 95% confidence intervals.

Figure 12 presents the coefficients of mental models of IS from the fully controlled specification. Relative to managers who predict 202, the standardized price ratio is 0.014 standard deviations higher for those who predict 192 (not statistically significant), 0.021 s.d. higher for those who predict 165 ($p < 0.05$), and 0.025 s.d. higher for those who predict 100 ($p < 0.05$). The estimates for managers who predict 165 and for managers who predict 100 are robust across all specifications. To put these numbers into context, predicting 165 rather than 202 is associated with a price difference comparable to facing one fewer competitor in the local market. Managers who predict 202 set lower prices than managers who predict 165 or 100, consistent with their higher perceived elasticities.

5.4 Mental models and profits

Managers who predict 202 charge lower prices than managers who predict 165 or 100 (Figure 12). What these lower prices imply for profit cannot be read off the correlation between price and profit, which could reflect reverse causality: managers might respond to low profits by lowering prices. We therefore estimate demand for our partner company's stations, using cost shifters as instruments for price, and use the estimates to evaluate the profit consequence of a counterfactual price change. The estimates give the change in producer surplus, which equals the change in profit.[7]

We model each station as facing a downward-sloping residual demand curve that includes both consumers buying less fuel and consumers switching to competing stations, so the elasticity already incorporates competitors' responses in the local market. Following a standard approach in the gasoline demand literature (Li et al. 2014; Coglianese et al. 2017), we assume constant elasticity of demand, specifying residual demand for station s as $q_s = A_s \cdot p_s^{-\varepsilon}$, where q_s is quantity sold, p_s is own price, $\varepsilon > 1$ is the price elasticity of demand, and A_s is a station-specific demand shifter.

To identify the elasticity, we exploit variation in marginal cost driven by the government-imposed price ceiling, which is indexed to world oil prices and plausibly uncorrelated with local demand shocks. Because our partner company is vertically integrated, under its internal accounting rules the transfer price that determines each station's marginal cost moves one-for-one with the ceiling. The ceiling is adjusted every ten working days. Instrumenting log price with the log of the ceiling, with station-product and time fixed effects, we obtain a pooled 2SLS estimate of $\hat{\varepsilon} = 1.65$. Appendix F.1 reports the estimating equation and the identifying variation.

With this elasticity, we calculate what the coefficients in Figure 12 imply for profit. The coefficients are in standard deviations of the price ratio, and we convert them into price gaps per liter. We take the average station as the baseline, with average price $\bar{p}$ and average sales volume $\bar{q}$, and back out the demand shifter as $\bar{A} = \bar{q} \cdot \bar{p}^{\,\hat{\varepsilon}}$. Marginal cost is observed directly in the internal transfer price. This avoids the common practice in empirical industrial organization of inferring marginal cost from prices under the assumption that firms price optimally (Berry et al. 1995; Nevo 2001), an assumption that is problematic when managers hold wrong models of demand. We set the higher price at the average price and the lower price at the average price minus the estimated price gap, and compare producer surplus, $(p - MC)\,q$, at the two prices. Appendix F.2 gives the steps. The price gaps are associations across managers conditional on the controls, so the calculation compares managers who hold different models rather than changing the model of any one manager.

Relative to managers who predict 165, the correct answer, the lower prices of managers

[7]The discrepancy between profit and producer surplus, namely fixed costs, is differenced out when calculating the change in profit.

who predict 202 imply 3.3% lower daily fuel profit at the average station. Relative to managers who predict 100, they imply 3.9% lower daily fuel profit. We do not report a profit gap for managers who predict 192, whose price gap is not significantly different from 202. Given the estimated demand, the lower prices of managers who predict 202 are a mistake from the perspective of the firm's profit. The calculation is static, comparing daily fuel profit at the two prices and leaving aside responses over longer horizons.

Robustness. As an alternative to constant elasticity, we also estimate linear demand, which separately identifies own-price and cross-price effects and incorporates competitor responses through conjectural variations (Slade 1986). The elasticity for the average station is then 1.63 in absolute value, and the same price gaps again imply 3.3% and 3.9% lower daily fuel profit. Appendix F.3 reports the specification and the estimates.

6 Conclusion

This paper provides evidence that a substantial fraction of firm managers neglect consumer intertemporal substitution when interpreting data on sales responses to price changes. Using field data from roughly 15,000 gas station managers, we document that about 40 percent of managers underestimate or neglect intertemporal substitution. This neglect is associated with overestimating the price elasticity of demand for fuel, and the profitability of price cuts. These managers charge significantly lower fuel prices in practice, consistent with failing to appreciate the inframarginal losses from discounts arising from consumer intertemporal substitution.

These findings have important implications. They identify a new factor that can explaining heterogeneity in pricing strategies. The point to a reason that average prices may tend to be lower than expected if firms were fully rational, which tends to shift surplus from firms to workers and improve market efficiency if it works against firms fully exploiting market power. Another implication is adding nuance to understanding a key puzzle in industrial organization, the reasons behind periodic discounts. While a leading explanation in the literature is that this reflects attempts to engage in dynamic price discrimination, our findings indicate that at least some managers may do this because they neglect consumer intertemporal substitution.

Regarding the puzzle of how this type of misperception can be persistent despite a data-rich environment, the paper provides a rich set of findings on mechanisms. One set of findings points to a feedback loop, between having a wrong model that the entire sales response to discounts is new customers, which directs attention to sales on discount days rather than dip days, which helps preserve the wrong model and the neglect of intertemporal substitution. Another finding is that low cognitive skills make individuals more likely to have the wrong model, less attentive to data patterns, less responsive to attention prompts, and less responsive to being assigned the correct model. This is consistent with lower cognitive skills

entailing higher costs of attention, which make it harder to notice the data patterns that reveal intertemporal substitution. Another result is that assigning people a version of the wrong model that is augmented with another factor that explains away the dips largely eliminates the impact of high cognitive skills on figuring out the correct model.

These findings on mechanisms also have important implications. They are in line with recent theoretical work in behavioral economics, providing supporting field evidence. They also have practical implications, indicating that firms seeking to improve managerial practices may need to do more than simply providing better data or analytics tools; firms may need to augment data with attention prompts, but also counteract a tendency to respond to noticing the patterns that wrong models deem irrelevant by augmenting those models with extra factors that improve fit but do not improve understanding.

Several directions for future research emerge from our work. First, while we document the prevalence and consequences of neglecting intertemporal substitution in retail fuel markets, it remains an open question how widespread this phenomenon is across other industries and product categories. Second, our evidence on the role of cognitive skills in shaping mental models and attention patterns suggests that manager selection and training practices could have important implications for firm performance and market outcomes. Finally, understanding whether and how managers update their mental models over longer time horizons, and what interventions might accelerate this learning, remains an important area for investigation. Our experimental evidence that attention prompts can causally reduce neglect of intertemporal substitution provides a promising starting point for designing such interventions in practice.

References

Ambuehl, Sandro and Heidi C. Thysen, "Choice with Competing Models: An Experimental Study," ECON Working Paper 458, Department of Economics, University of Zurich 2025. Revised November 2025.

_ , Rahul Bhui, and Heidi C. Thysen, "Mental Models of Causal Structure in Economics and Psychology," 2026. arXiv:2603.29070, April 2026.

Anderson, Richard C. and James W. Pichert, "Recall of Previously Unrecallable Information Following a Shift in Perspective," *Journal of Verbal Learning and Verbal Behavior*, 1978, *17* (1), 1–12.

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart, "Subjective Models of the Macroeconomy: Evidence From Experts and Representative Samples," *The Review of Economic Studies*, 2022, *89* (6), 2958–2991.

_ , **Ingar Haaland, Christopher Roth, Mirko Wiederholt, and Johannes Wohlfart**, "Narratives about the Macroeconomy," *The Review of Economic Studies*, 2026. Advance access, published online 18 February 2026.

_ , **Philipp Schirmer, and Johannes Wohlfart**, "Mental Models of the Stock Market," *The Quarterly Journal of Economics*, 2026. Advance access, published online 25 July 2026.

Ba, Cuimin, "Robust Misspecified Models," *American Economic Review*, 2026, *116* (4), 1340–1379.

Barron, Kai and Tilman Fries, "Narrative Persuasion," Discussion Paper SP II 2023–301r, WZB Berlin Social Science Center 2024. January 2023, revised April 2024.

Bénabou, Roland and Jean Tirole, "Mindful economics: The production, consumption, and value of beliefs," *Journal of Economic Perspectives*, 2016, *30* (3), 141–64.

Berk, Robert H., "Limiting Behavior of Posterior Distributions when the Model is Incorrect," *The Annals of Mathematical Statistics*, 1966, *37* (1), 51–58.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

Bilker, Warren B, John A Hansen, Colleen M Brensinger, Jan Richard, Raquel E Gur, and Ruben C Gur, "Development of abbreviated nine-item forms of the Raven's standard progressive matrices test," *Assessment*, 2012, *19* (3), 354–369.

Bohren, J. Aislinn and Daniel N. Hauser, "Misspecified Models in Learning and Games," *Annual Review of Economics*, 2025, *17*, 427–451.

Boizot, Christine, Jean-Marc Robin, and Michael Visser, "The demand for food products: an analysis of interpurchase times and purchased quantities," *The Economic Journal*, 2001, *111* (470), 391–419.

Brewer, William F. and James C. Treyens, "Role of Schemata in Memory for Places," *Cognitive Psychology*, 1981, *13* (2), 207–230.

Charles, Constantin and Chad Kendall, "Causal Narratives," 2025. Working paper, November 2025. Earlier version circulated as NBER Working Paper 30346 (2022).

Cho, In-Koo and Kenneth Kasa, "Learning and Model Validation," *The Review of Economic Studies*, 2015, *82* (1), 45–82.

Coglianese, John, Lucas W Davis, Lutz Kilian, and James H Stock, "Anticipation, tax avoidance, and the price elasticity of gasoline demand," *Journal of Applied Econometrics*, 2017, *32* (1), 1–15.

DellaVigna, Stefano and Matthew Gentzkow, "Uniform pricing in US retail chains," *The Quarterly Journal of Economics*, 2019, *134* (4), 2011–2084.

Doherty, Michael E., Clifford R. Mynatt, Ryan D. Tweney, and Michael D. Schiavo, "Pseudodiagnosticity," *Acta Psychologica*, 1979, *43* (2), 111–121.

Drew, Trafton, Melissa L.-H. Võ, and Jeremy M. Wolfe, "The Invisible Gorilla Strikes Again: Sustained Inattentional Blindness in Expert Observers," *Psychological Science*, 2013, *24* (9), 1848–1853.

Ellison, Glenn, "Bounded rationality in industrial organization," *Econometric Society Monographs*, 2006, *42*, 142.

Enke, Benjamin, "What You See Is All There Is," *The Quarterly Journal of Economics*, 2020, *135* (3), 1363–1398.

_ **and Florian Zimmermann**, "Correlation Neglect in Belief Formation," *The Review of Economic Studies*, 2019, *86* (1), 313–332.

_ **and Thomas Graeber**, "Cognitive uncertainty," *The Quarterly Journal of Economics*, 2023, *138* (4), 2021–2067.

Esponda, Ignacio and Demian Pouzo, "Berk–Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models," *Econometrica*, 2016, *84* (3), 1093–1130.

_ **and Emanuel Vespa**, "Endogenous Sample Selection: A Laboratory Study," *Quantitative Economics*, 2018, *9* (1), 183–216.

_ , _ **, and Sevgi Yuksel**, "Mental Models and Transfer Learning," *AEA Papers and Proceedings*, 2023, *113*, 659–664.

_ , _ **, and** _ , "Mental Models and Learning: The Case of Base-Rate Neglect," *American Economic Review*, 2024, *114* (3), 752–782.

Fréchette, Guillaume R., Emanuel Vespa, and Sevgi Yuksel, "Extracting Models From Data Sets: An Experiment," 2025. Working paper, version of October 28, 2025, SSRN 5026751.

Frick, Mira, Ryota Iijima, and Yuhta Ishii, "Welfare Comparisons for Biased Learning," *American Economic Review*, 2024, *114* (6), 1612–1649.

Fudenberg, Drew and Giacomo Lanzani, "Which Misspecifications Persist?," *Theoretical Economics*, 2023, *18* (3), 1271–1315.

_ , _ **, and Philipp Strack**, "Limit Points of Endogenous Misspecified Learning," *Econometrica*, 2021, *89* (3), 1065–1098.

Gabaix, Xavier, "A sparsity-based model of bounded rationality," *The Quarterly Journal of Economics*, 2014, *129* (4), 1661–1710.

_ , "Behavioral inattention," in "Handbook of behavioral economics: Applications and foundations 1," Vol. 2, Elsevier, 2019, pp. 261–343.

Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein, "Channeled Attention and Stable Errors," August 2026. Working paper.

Goldfarb, Avi and Mo Xiao, "Who thinks about the competition? Managerial ability and strategic entry in US local telephone markets," *American Economic Review*, 2011, *101* (7), 3130–3161.

_ **and** _ , "Transitory Shocks, Limited Attention, and a Firm's Decision to Exit," *Quantitative Marketing and Economics*, 2024, *22* (3), 223–255.

Golman, Russell, David Hagmann, and George Loewenstein, "Information avoidance," *Journal of economic literature*, 2017, *55* (1), 96–135.

Graeber, Thomas, "Inattentive Inference," *Journal of the European Economic Association*, 2023, *21* (2), 560–592.

Grass, Paul, Philipp Schirmer, and Malin Siemers, "Sticky Models," ECONtribute Discussion Paper 355, University of Bonn and University of Cologne 2026. Updated version dated June 19, 2026; also CRC TR 224 Discussion Paper No. 763; SSRN 5384226.

Han, Yi, David Huffman, and Yiming Liu, "Minds, models and markets: How managerial cognition affects pricing strategies," Technical Report, Working Paper 2025a.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein, "Learning Through Noticing: Theory and Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 2014, *129* (3), 1311–1353.

He, Kevin and Jonathan Libgober, "Misspecified Learning and Evolutionary Stability," *Journal of Economic Theory*, 2025, *230*, 106082.

Heidhues, Paul and Botond Kőszegi, "Behavioral industrial organization," *Handbook of Behavioral Economics: Applications and Foundations 1*, 2018, *1*, 517–612.

_ **, Botond Kőszegi, and Philipp Strack**, "Unrealistic Expectations and Misguided Learning," *Econometrica*, 2018, *86* (4), 1159–1214.

_ **,** _ **, and** _ , "Convergence in Models of Misspecified Learning," *Theoretical Economics*, 2021, *16* (1), 73–99.

Hendel, Igal and Aviv Nevo, "Sales and consumer inventory," *The RAND Journal of Economics*, 2006a, *37* (3), 543–561.

_ and _ , "Measuring the implications of sales and consumer inventory behavior," *Econometrica*, 2006b, *74* (6), 1637–1673.

Hong, Harrison, Jeremy C. Stein, and Jialin Yu, "Simple Forecasts and Paradigm Shifts," *The Journal of Finance*, 2007, *62* (3), 1207–1242.

Hortaçsu, Ali and Steven L Puller, "Understanding strategic bidding in multi-unit auctions: a case study of the Texas electricity spot market," *The RAND Journal of Economics*, 2008, *39* (1), 86–114.

_ , Fernando Luco, Steven L Puller, and Dongni Zhu, "Does strategic ability affect efficiency? Evidence from electricity markets," *American Economic Review*, 2019, *109* (12), 4302–4342.

Jenkins, Herbert M. and William C. Ward, "Judgment of Contingency Between Responses and Outcomes," *Psychological Monographs: General and Applied*, 1965, *79* (1), 1–17. Whole No. 594.

Kendall, Chad and Ryan Oprea, "On the Complexity of Forming Mental Models," *Quantitative Economics*, 2024, *15* (1), 175–211.

Klayman, Joshua and Young-Won Ha, "Confirmation, Disconfirmation, and Information in Hypothesis Testing," *Psychological Review*, 1987, *94* (2), 211–228.

Lanzani, Giacomo, "Dynamic Concern for Misspecification," *Econometrica*, 2025, *93* (4), 1333–1370.

Levy, Gilat, Ronny Razin, and Alwyn Young, "Misspecified Politics and the Recurrence of Populism," *American Economic Review*, 2022, *112* (3), 928–962.

Li, Shanjun, Joshua Linn, and Erich Muehlegger, "Gasoline Taxes and Consumer Behavior," *American Economic Journal: Economic Policy*, 2014, *6* (4), 302–342.

List, John A and Charles F Mason, "Are CEOs expected utility maximizers?," *Journal of Econometrics*, 2011, *162* (1), 114–123.

_ , Ian Muir, Devin Pope, and Gregory Sun, "Left-digit bias at lyft," *Review of Economic Studies*, 2023, *90* (6), 3186–3237.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, "Rational Inattention: A Review," *Journal of Economic Literature*, 2023, *61* (1), 226–273.

Nevo, Aviv, "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 2001, *69* (2), 307–342.

Olea, José Luis Montiel, Pietro Ortoleva, Mallesh M. Pai, and Andrea Prat, "Competing Models," *The Quarterly Journal of Economics*, 2022, *137* (4), 2419–2457.

Ortoleva, Pietro, "Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News," *American Economic Review*, 2012, *102* (6), 2410–2436.

Pesendorfer, Martin, "Retail sales: A study of pricing behavior in supermarkets," *The Journal of Business*, 2002, *75* (1), 33–66.

Schwartzstein, Joshua and Adi Sunderam, "Using Models to Persuade," *American Economic Review*, 2021, *111* (1), 276–323.

Simons, Daniel J. and Christopher F. Chabris, "Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events," *Perception*, 1999, *28* (9), 1059–1074.

Sims, Christopher A., "Implications of Rational Inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665–690.

Slade, Margaret E, "Conjectures, firm characteristics, and market structure," *International Journal of Industrial Organization*, 1986, *4* (4), 347–369.

Smedslund, Jan, "The Concept of Correlation in Adults," *Scandinavian Journal of Psychology*, 1963, *4* (1), 165–173.

Spiegler, Ran, "Bayesian Networks and Boundedly Rational Expectations," *The Quarterly Journal of Economics*, 2016, *131* (3), 1243–1290.

Strulov-Shlain, Avner, "More than a penny's worth: Left-digit bias and firm pricing," *Review of Economic Studies*, 2023, *90* (5), 2612–2645.

Tadelis, Steven, Christopher Hooton, Utsav Manjeer, Daniel Deisenroth, Nils Wernerfelt, Nick Dadson, and Lindsay Greenbaum, "Learning, Sophistication, and the Returns to Advertising: Implications for Differences in Firm Performance," Technical Report, National Bureau of Economic Research 2023.

Wason, Peter C., "On the Failure to Eliminate Hypotheses in a Conceptual Task," *Quarterly Journal of Experimental Psychology*, 1960, *12* (3), 129–140.

Zimmermann, Florian, "The dynamics of motivated beliefs," *American Economic Review*, 2020, *110* (2), 337–61.

Online Appendix

A Additional results on the conceptual framework

A.1 Proofs

Throughout we condition on the realized sequence of prices, which the manager sees. All probability statements are over sales given that sequence. The outcome is sales. Write π^* for the truth, π_C for the correct model and π_S for the simple wrong model. A candidate of π_S is a pair (H, L), the sales value at the regular price and the sales value at the discount price. Candidates are indexed by integer pairs and the prior assigns positive mass to each. The lattice matters, since under an atomless prior every single value the manager reads would have probability zero and the model would be refuted at once. Write β_H for the mass on the regular-price value he reads and β for the joint mass on the pair he ends up holding, with $0 < \beta \leq \beta_H \leq 1$. The correct model's prior places positive mass on every integer pair (D, δ) that satisfies the two restrictions, the true pair (D^*, δ^*) among them. The data noticed by t are n^t, and $P(n^t \mid \pi)$ is the probability a model π assigns to them, averaging over its candidates by its prior. The manager's task is the one stated in the text, to work out the demand curve: he says what daily sales would be at each of the two prices in the table, were that price permanent, and any error in either value lowers his payoff. A noticing strategy is a sequence of partitions of histories, as in Gagnon-Bartsch, Rabin and Schwartzstein (2026). Their attentional strategy pairs it with a rule that maps what is noticed to an answer. Because the task does not depend on the history (their Assumption 2), their sufficiency (their Definition 1) is a condition on the noticing strategy alone.

Proof of Prediction 1. Let τ be the first regular-price day the manager reads, the first day of the table if he reads in calendar order. Consider the noticing strategy that notices the sales value on day τ and on the first discount day he reads and nothing else about sales. Because the announced schedule fixes the price on every day and all probabilities condition on it, the strategy draws no distinction by price path.

This is the strategy Assumption 1 prescribes for the simple wrong model. Every candidate assigns all regular-price days one value and all discount days another, each with probability one. Before the manager has read a value at a price, the candidates with positive posterior mass allow that value to differ, and he notices it. Once he has read a value at a price, every candidate with positive posterior mass assigns that value to every day at that price, and he notices nothing further about sales at that price. Once he has read one value at each price the posterior collapses to a single candidate.

The strategy is sufficient for the manager's task: under the simple wrong model sales at a permanent price are the value at that price, the only unknowns are H and L, and the

two noticed values fix both. It is also minimal in the sense of Gagnon-Bartsch, Rabin and Schwartzstein (2026, Definition 2): it draws only two distinctions, by the value on day τ and by the value on the first discount day he reads. A coarser strategy that merged histories differing in either value would leave positive posterior mass on more than one value of H or of L, and with it uncertainty about sales at that price that lowers his expected payoff. The strategy never averages regular-price days: after day τ it notices nothing about their sales. When τ is a baseline day, as the first day of the table is, the values 90, 82 and 91 never enter the noticed data, and never move his belief about H.

Along the true path the noticed data consist of the value v_τ on day τ and the value 202 on the first discount day he reads. Both values have probability one under the truth, because sales are deterministic given the price path. Under π_S the noticed data therefore have probability one until he reads a value, the marginal prior mass on the first value he reads (β_H when that value is v_τ) until he has read a value at each price, and β from then on, while $P(n^t \mid \pi^*) = 1$ throughout. Each of these is at least β, and the likelihood ratio is bounded below by $\beta > 0$ at every t: it never reaches zero in finite time and its limit inferior is β. No noticed data are ever impossible under the simple wrong model, and by Assumption 1 the manager never gives it up.

The same bound gives stability in the sense of Gagnon-Bartsch, Rabin and Schwartzstein (2026, Definition 3) with respect to any alternative, which they call a lightbulb model, not only with respect to the truth. Outcomes are discrete: for any prior λ over candidate processes $P(n^t \mid \lambda) \leq 1$, and therefore $P(n^t \mid \pi_S)/P(n^t \mid \lambda) \geq \beta$ for every λ and every t. What does the work is that $P(n^t \mid \pi_S)$ is bounded away from zero absolutely, not that the truth fits the noticed data perfectly. □

The main text calls the simple wrong model censored, in the sense of Gagnon-Bartsch, Rabin and Schwartzstein (2026, Definition B.1), once the manager has read the table. Their definition is stated for the prior, and we apply it to his posterior. We prove stability directly rather than citing their result for censored models (their Proposition B.2), which requires, among other conditions, a stationary environment and candidates that assign every outcome in the model's support a probability strictly between zero and one. Our schedule is not stationary, our candidates are deterministic, and the argument above uses only the lower bound β.

Proof of Prediction 2. The manager's perceived demand pairs are $(5.9, v_\tau)$ and $(5.2, 202)$, where $v_\tau \in \{100, 90, 82, 91\}$ is the value he read; the truth, at permanent prices, gives $(5.9, 100)$ and $(5.2, 165)$. Since $v_\tau \leq 100$, the perceived quantity response to the price cut, $202 - v_\tau \geq 102$, exceeds the true response of 65. The arc elasticity between the two prices is the midpoint quantity change divided by the midpoint price change, which equals 0.1261 in both cases. Perceived, the quantity change is $2(202 - v_\tau)/(202 + v_\tau)$, which is decreasing in v_τ and equals 0.6755 at $v_\tau = 100$, compared with 0.4906 for the truth. The elasticities are

therefore about 5.4 and 3.9, and the perceived one exceeds the true one for every $v_\tau \le 100$. No functional form is attributed to the manager, and he is not asked to extrapolate, since the counterfactual price is one of the two prices in the table.

For pricing, let $c < 5.2$ denote unit cost. The perceived profit change from a permanent cut to 5.2 exceeds the true change by

$$(5.2-c)\big(202-165\big) \;+\; (5.9-c)\big(100-v_\tau\big) \;\ge\; (5.2-c)\,\delta^*(0) \;>\; 0.$$

Whenever the true change is negative and the perceived change positive, an otherwise rational manager cuts the price and earns less than at the regular price. With $v_\tau = 100$ the true change is $268-65c$ and the perceived change is $460.4-102c$. This holds for any unit cost between 268/65 and 460.4/102, that is between 4.13 and 4.51, an interval that contains the cost of 4.3 implied by the profit column of Table 1. The comparison is against a permanent regular price, not against the retailer's own policy of weekly discounts. Against a richer menu of prices the same conclusion holds under a constant-elasticity demand curve, where the perceived optimum $c\,\hat{\varepsilon}/(\hat{\varepsilon}-1)$ is decreasing in the perceived elasticity $\hat{\varepsilon}$ and so lies below the true optimum. □

Proof of Prediction 3. Fix any noticing strategy that separately notices the sales values of two regular-price days whose values differ, say 100 and 82. Under every candidate (H, L) of the simple wrong model, sales on a regular-price day equal H with probability one. Noticed data containing two regular-price days with different values therefore have probability zero under every candidate, and hence under the model for any prior or posterior over (H, L). By Assumption 1 the manager gives up the simple wrong model. Each window our attention prompts name, Days 4–7 and Days 10–13, contains the values 90, 82 and 91. A strategy that notices the values in either window satisfies the hypothesis on its own. □

The simple wrong model fails to conceptualize an event in the sense of Gagnon-Bartsch, Rabin and Schwartzstein (2026, Definition 5): it assigns probability zero to any history containing two regular-price days with different sales, and the truth produces such histories. By their Proposition 2, everything a person notices under a minimal sufficient attentional strategy has positive probability under his model, and his own noticing never exposes the failure. By their Proposition 9, a model that fails to conceptualize an event is fragile to temporary-noticing refinements (their Definitions 11 and 12). Our attention prompts are such a refinement, of the kind they read as a stimulus impossible to ignore, and the proof above shows that the prompts dislodge the simple wrong model.

Proof of Prediction 4. Entertaining the alternative with weight $\alpha \in (0, 1)$ forms the mixture $(1-\alpha)\,\pi + \alpha\,\pi'$, the perturbed model of Gagnon-Bartsch, Rabin and Schwartzstein (2026, Definition 13), and by the extension of Assumption 1 the manager notices any dis-

tinction across which either component, given what he has noticed, allows sales to differ.

(i) Let $\pi = \pi_S$ and $\pi' = \pi_C$. Under π_C sales on a regular-price day depend on its position relative to the discount through δ. The mixture therefore allows the values on the dip days to differ from the value on baseline days, and the manager notices those values together with their positions. With the table available the pairing is assembled from the page at no cost. The assembled data contain regular-price days with distinct values, 100 on baseline days and 90, 82 and 91 on the dip days, an event with probability zero under every candidate of π_S and probability one under the correct-model candidate with the true parameters. The first update on the assembled data therefore places zero posterior mass on every candidate of the simple wrong model.

(ii) Let $\pi = \pi_C$ and $\pi' = \pi_S$. Under Assumption 1 the manager has already noticed the values on the dip days, since the correct model allows them to differ from the value on baseline days. At the encounter the data he has noticed contain those values as distinct values at the regular price. Those data have probability zero under every candidate of π_S. The first update removes the entertained weight α entirely and returns the posterior over correct-model candidates to what it was. The step relies on Assumption 1 having the manager notice each of the three values around a discount, which is more than the task requires. Sales at a permanent 5.2 depend on the discount-day value and the three values around it only through their sum, and data showing only the value 100 and the sum 465 would have probability one under the candidate of the simple wrong model that pairs 100 with 165. □

Gagnon-Bartsch, Rabin and Schwartzstein (2026) read the perturbed model as permanent adoption of the alternative, where we have in mind a passing thought. Their results on perturbations concern whether the perturbed model admits a stable attentional strategy. Which component the posterior concentrates on, the content of Prediction 4, is not part of their analysis.

Two remarks on the empirical counterparts. The predictions are stated for the model a manager holds. In the experiments we assign a model rather than observe the one held. The contrasts we report are therefore between assigned conditions rather than between holders. Elicited confidence is an imperfect proxy for whether a manager entertains an alternative. It is an inverse one, on the conjecture that a manager certain of his model sees nothing left to learn from an alternative, and it is imperfect on two counts: the predictions turn only on whether α is positive, not on its size, which leaves confidence to proxy the extensive margin alone; and what we elicit is confidence in the model and confidence in the prediction jointly, of which only the first belongs in the role α plays. Making α a function of how dogmatically the incumbent model is held would extend the framework rather than apply it.

Profits and the manager's own data. A manager pricing his own station differs from a subject reading the table in two ways: he earns the lower profit of Prediction 2, and he pro-

duces the data he sees. In the framework the lower profit plays no role in whether he gives up the simple wrong model. Bohren and Hauser (2025) group criteria for revising or ranking models into those based on poor fit with the data (Cho and Kasa, 2015; Montiel Olea et al., 2022; Lanzani, 2025; Ba, 2026) and those based on unexpectedly low payoffs (Levy et al., 2022; Fudenberg and Lanzani, 2023; Frick et al., 2024; He and Libgober, 2025). The second sentence of Assumption 1 is a criterion of the first kind, applied to the noticed data. Gagnon-Bartsch, Rabin and Schwartzstein (2026)'s Proposition 6 shows that for any wrong model some choice environment admits an arbitrarily costly stable attentional strategy: whether a person discovers an error is not closely tied to its cost. The framework also takes the data to be independent of the decision maker's actions, as their Assumption 1 does. The table satisfies this by construction, since the manager did not generate it. The data a station manager sees do not, because his own schedule is part of the reason he never sees a permanent discount. With data the decision maker does not produce, only correctly specified models are robust under the full-data switching rule of Ba (2026, footnote 19). With data he produces, a wrong model can persist because the data it leads him to generate fit it at a self-confirming equilibrium, as in Cho and Kasa (2015), without any role for attention. We therefore read the field evidence as consistent with the framework rather than as a test of it.

A.2 The predictor-neglect model

The predictor-neglect model of the main text conditions on the day but not on a day's position relative to the discount. Its candidates are

$$q_t \;=\; D(p_t) + b_t,$$

where $\{b_t\}$ is a deterministic profile of day-specific demand, unrestricted across days, and a property of the calendar rather than of the price path. Nothing is left to noise, and nothing about the profile is tied to where the discounts fall. The table identifies D and $\{b_t\}$ only up to an additive constant, and it leaves the split of the discount-day value between $D(5.2)$ and the calendar undetermined. The manager's own reading supplies both. He takes the baseline-day value, 100, as $D(5.9)$, and 202 as $D(5.2)$. The calendar effect is then zero on baseline days and on the two dates on which a discount happened to fall. The fitted effects are then -10 on Days 4 and 10, -18 on Days 6 and 12, and -9 on Days 7 and 13.

Claim 1 (Immunity of the predictor-neglect model)**.** *Under a prior with full support over integer day effects, noticed data on sales values have positive probability under the model, however finely the manager's attention is directed; no noticing strategy defined over sales values refutes it. Under the reading above its answer to the counterfactual is 202 on any date with no calendar effect, which is the answer that neglects intertemporal substitution, and* $202 + \bar{b}$

averaged over a week, where $\bar{b} = -37/7$, or about 196.7.

Proof of Claim 1. For any realized profile $\{v_t\}$ of regular-price sales values, the candidate with $b_t = v_t - D(5.9)$ assigns the data probability one. Under a full-support prior over the fourteen integer day effects the noticed data therefore have positive probability whatever distinctions attention is directed to. The counterfactual statement follows from reading the discount-day value as $D(5.2) = 202$ and the day effects as invariant to the price path: at a permanent 5.2, sales on date t are $202 + b_t$; averaging the fitted effects -10, -18 and -9 over the seven days of a week gives $202 - 37/7 \approx 196.7$. □

The average is not identified independently of the reading, and a manager who took $D(5.9)$ to be something other than 100 would land elsewhere. The model supplies no intertemporal substitution whatever the reading, and its answer on a date with no calendar effect therefore sits at the discount-day value rather than below it.

Three remarks bound the claim. First, disciplined special cases of the model are refutable. A pattern that repeats by day of the week cannot fit Table 1: the two discounts fall on different days of their respective weeks, the fifth and the fourth, and the dips follow the discount: the third day of the week is a baseline day in one week and low in the other, the sixth is low in both weeks but at different values, and the seventh is low in one and a baseline day in the other. A subject reaching for weekly seasonality is contradicted once he notices the values. It is the unrestricted day-specific version that fits everything. It supplies no intertemporal substitution, and gives the wrong counterfactual. Second, what would separate the model from the truth is variation in the schedule, since moving the discount relocates the dips while calendar effects stay put. Fourteen days, in which each date occurs once, do not supply that variation, because any profile of regular-price values is fitted exactly by some profile of day effects, and a manager can suppose that two more weeks would repeat the pattern he has seen. In the terms of Gagnon-Bartsch, Rabin and Schwartzstein (2026), the model is not identifiably wrong on this schedule (their Definition 4), and unlike the simple wrong model it conceptualizes all possible events (their Definition 5). Our design separates the two models instead by supplying the correct model. Third, idiosyncratic noise would accommodate the values equally well, but the instructions foreclose it: managers are told that the table shows all the variables that may affect profit, subjects that it shows all the variables that can affect sales, and both that the data are representative of all possible situations. Deterministic day effects are the accommodation that the instructions leave open. They, and not noise, are the relevant caveat to Prediction 3.

The simple wrong model and the predictor-neglect model leave different traces in what a manager can report about the table, which is the basis for the conjecture in the main text. A holder of the simple wrong model has one regular-price value and a partition. He should report that value on every regular-price day, or report it once and decline the rest; two different regular-price values should not appear. A holder of the predictor-neglect model

has noticed the values and should report low ones. He has not noticed what predicts them, and should not place them reliably at the dip days. Directing attention should therefore cut the share reporting the regular-price value everywhere while leaving roughly unchanged the share reporting a low value at the wrong day. The framework does not deliver this, since it fixes neither the mix of the two models in a population nor how a manager reports a value he did not notice.

B Related literature from psychology

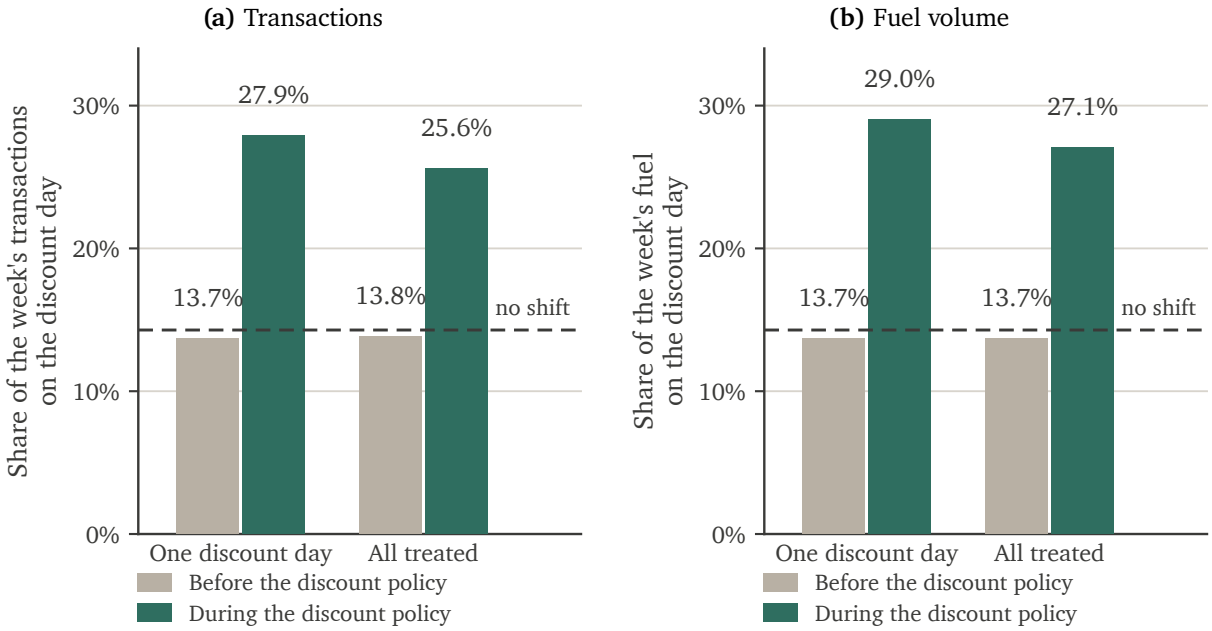
Our paper complements a body of research in psychology showing in various ways that what a person is doing determines what the person sees and remembers (Anderson and Pichert 1978; Brewer and Treyens 1981; Simons and Chabris 1999; Drew, Võ, and Wolfe 2013). In the canonical demonstration of inattentional blindness, viewers who have been instructed to count the passes made by one of two intermingled basketball teams, and who must track and filter a moving display to do so, fail to report a person in a gorilla suit walking slowly through the middle of it (Simons and Chabris 1999). In the corresponding demonstration for memory, readers of a story about two boys spending the day at one boy's house recall different details of the same text depending on whether they were asked to read it from the point of view of a burglar or of a prospective homebuyer, and they produce details they had omitted the first time when the perspective is switched and they are asked to recall it again (Anderson and Pichert 1978).

It also complements research demonstrating systematic biases in how humans verify the support for causal relationships using data (Smedslund 1963; Jenkins and Ward 1965; Doherty et al. 1979; Klayman and Ha 1987). In demonstrations of covariation neglect, subjects who are handed complete data on cases—for each patient, whether a symptom was present and whether a diagnosis was present—judge whether the two are related largely on the strength of the cases in which both appear, giving little weight to the cases that bear equally on the question, so that their judgments end up nearly uncorrelated with the true contingency (Smedslund 1963). In one canonical demonstration of the positive test strategy, by contrast, subjects must discover a rule the experimenter has in mind, and the only evidence available is evidence they generate themselves: they submit triples of numbers of their own choosing and learn of each only whether it satisfies the rule. They almost always submit triples that their current hypothesis says should qualify, rather than ones that would expose it as wrong: having inferred "ascending by twos" from the starting example 2, 4, 6, they try 8, 10, 12 instead of 3, 9, 11. Because the rule is merely "ascending," the data they assemble consist entirely of confirmations, and they announce the wrong rule with confidence (Wason 1960; Klayman and Ha 1987).

What this literature has not established is that a belief about how the world works, assigned from outside, determines which parts of complete and freely inspectable statistical data get noticed—and that the resulting error persists, to some extent resists correction by encountering the correct model, and has consequences beyond judgment for actual economic decisions.

C Evidence on the existence of intertemporal substitution in fuel markets

Figure C.1: Share of regular customers' weekly transactions and fuel volume on the discount day



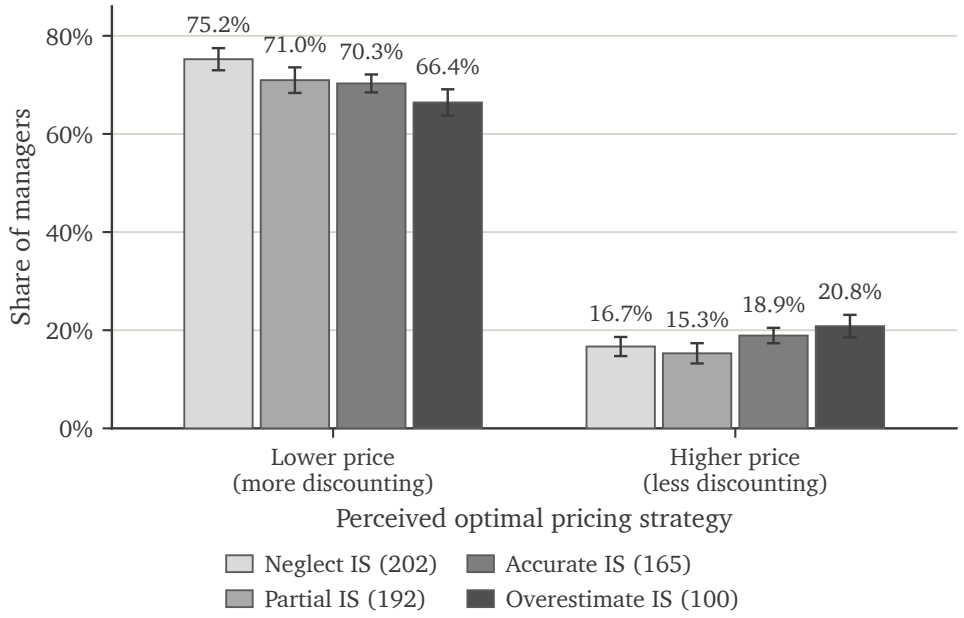
Notes: This figure shows the share of regular customers' weekly transactions and fuel volume on the discount day, before the policy starts and during the policy. Panel (a) shows transactions and Panel (b) shows fuel volume. Regular customers are loyalty card holders with at least two transactions in the month before a station's weekly discount policy starts and at least four transactions in the two months before. The treatment is a regularly occurring weekly price discount on a given day of the week. Some stations have only one discount day per week (usually Saturday); this group is denoted one discount day stations. Some other stations have two discount days per week (usually Saturday and Friday); the all treated stations group includes these along with the one discount day stations. For each week the share is divided by the number of discount days active that week, so the share with no shift in timing is one day in seven (14.3%) at every station, marked by the dashed line. The sample includes 10,629 customer-station pairs at 173 stations; the one discount day subsample includes 5,053 customer-station pairs at 85 stations.

D Additional results on predictions in the detergent task

Figure D.2 shows views on optimal pricing by all four predictions in the detergent task. The results reveal a monotonic relationship between the prediction in the detergent task and pricing beliefs. The proportion of managers choosing the lower price strategy (more discounting) decreases systematically across the four predictions: 75.2% among managers who predict 202, 73.1% among managers who predict 192, 69.8% among managers who predict 165, and only 65.4% among managers who predict 100. Conversely, the proportion choosing the higher price strategy (less discounting) increases monotonically in the same order.

The monotonic pattern is consistent with our hypothesized mechanism: managers who neglect intertemporal substitution systematically overestimate demand elasticity. The better managers understand that temporary discounts cause customers to shift purchase timing rather than expand total demand, the more conservative they become about the profitability of aggressive discounting. Managers who predict 165 or 100 are most likely to favor maintaining higher prices, understanding that much of the sales boost from discounts is illusory—representing intertemporal substitution rather than gained new demand.

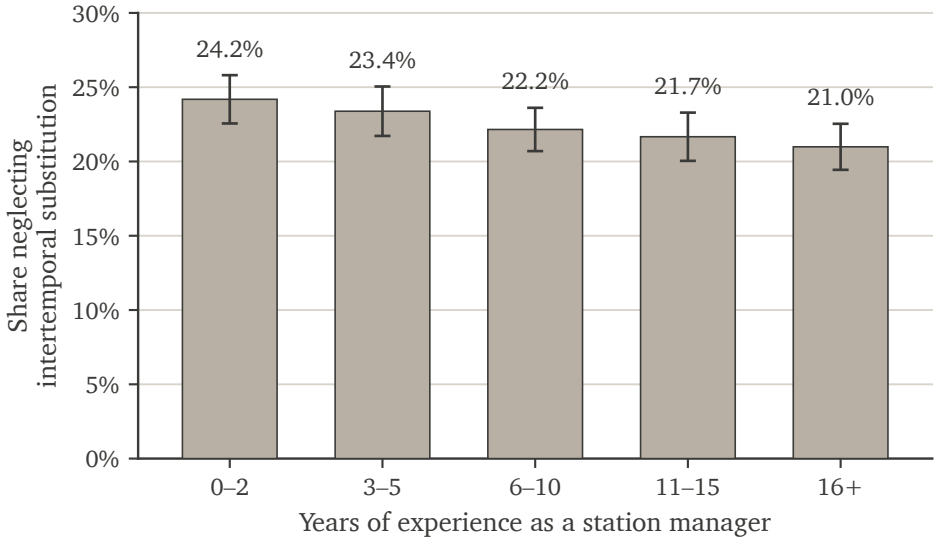
Figure D.2: Mental models of intertemporal substitution and perceived optimal price: four categories



Notes: This figure shows the share of managers choosing each perceived optimal pricing strategy, by their prediction in the detergent task. The lower price strategy involves more discounting and the higher price strategy less discounting. The sample includes the $N = 6{,}164$ managers who answered the pricing question. Error bars show 95% confidence intervals.

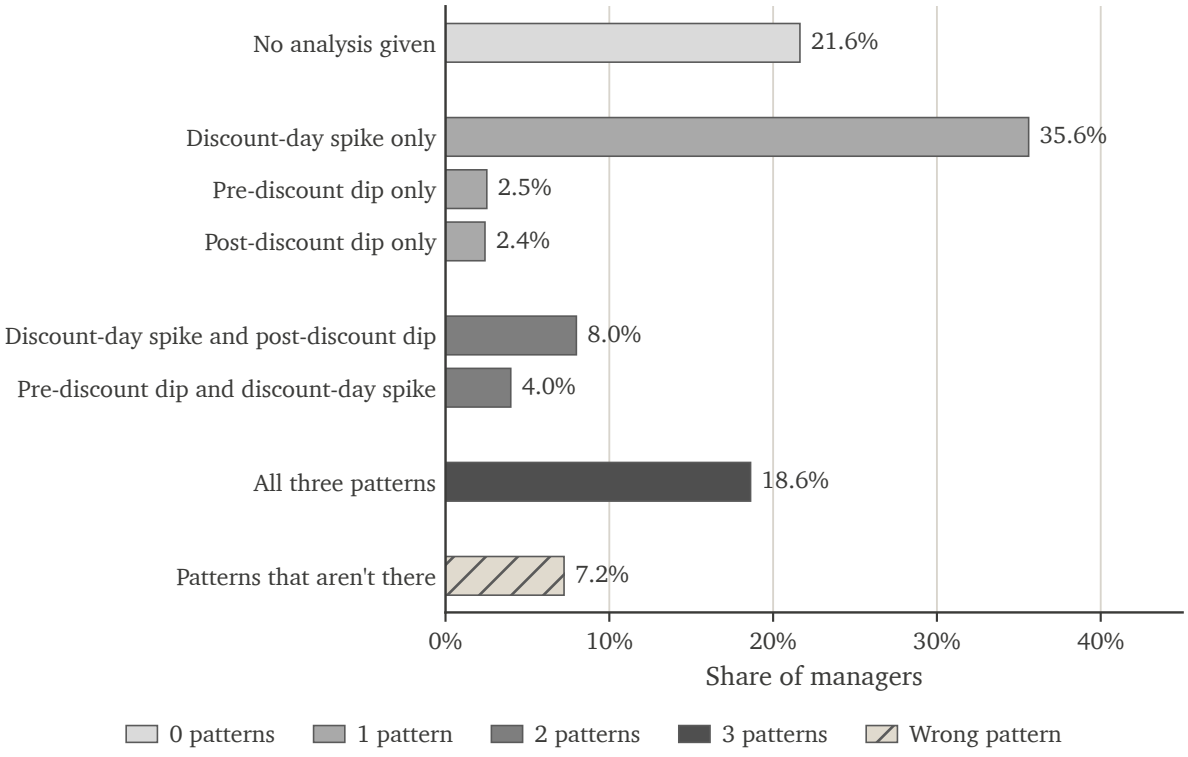
E Additional results on mechanisms underlying manager neglect of intertemporal substitution

Figure D.3: Neglect of intertemporal substitution and experience as a station manager



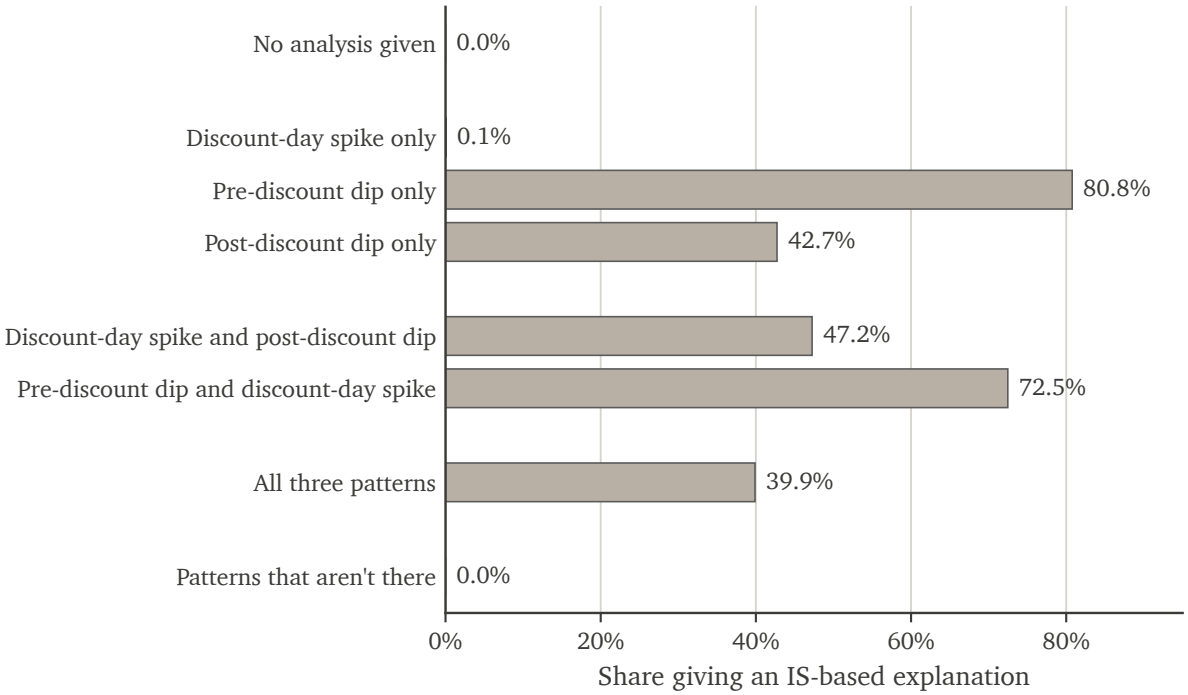
Notes: This figure shows the share of managers who predict 202 in the detergent task, the answer that neglects intertemporal substitution, by years of experience as a station manager. Error bars show 95% confidence intervals.

Figure E.4: Type and number of data patterns mentioned in manager explanations



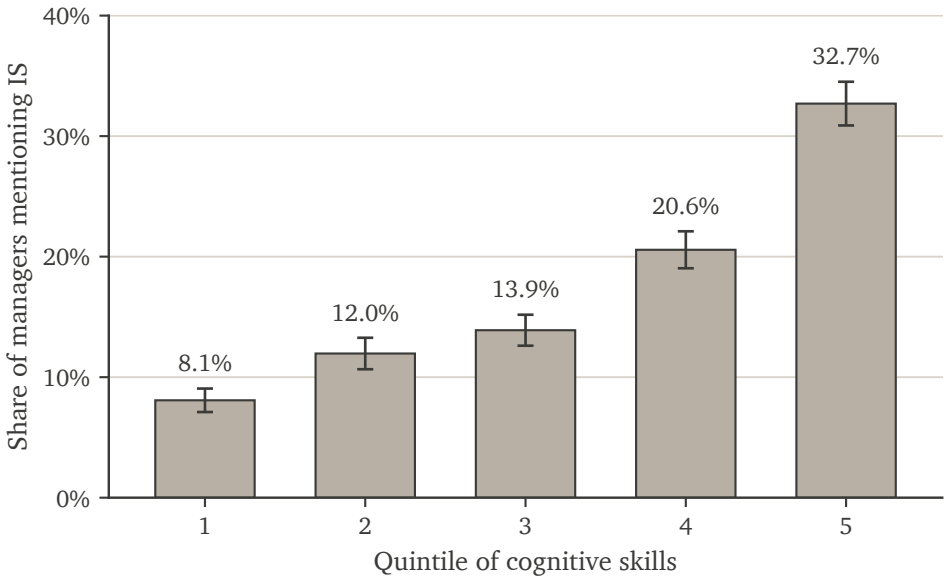
Notes: This figure shows the share of managers whose explanation of the detergent data mentions each combination of the three data patterns: the pre-discount dip, the discount-day spike and the post-discount dip. Shading gives the number of patterns mentioned, and the bars are grouped by that number. Explanations were classified by a large language model against a rubric developed on a random sample of 1,000 responses. The sample includes $N = 13{,}407$ managers, and the categories partition the sample.

Figure E.5: Frequencies of explanation based on intertemporal substitution, conditional on number and type of patterns mentioned



Notes: This figure shows the share of managers whose explanation of the detergent data involves intertemporal substitution, by which of the three data patterns the explanation mentions. The categories are those of Figure E.4. Shares of zero are exact: no manager who gave no analysis, and none who mentioned only patterns that are not there, gave an explanation based on intertemporal substitution. The sample includes $N = 13{,}407$ managers.

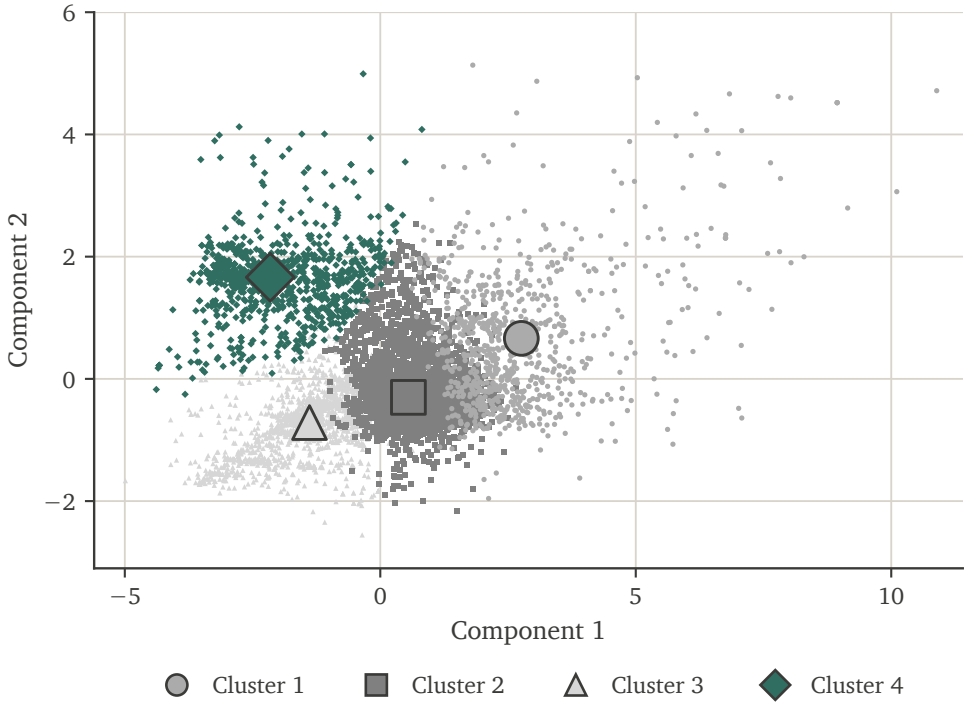
Figure E.6: Frequency of mentioning intertemporal substitution in explanation of data by cognitive skills



Notes: This figure shows the share of managers whose explanation of the detergent data mentions intertemporal substitution, by quintile of cognitive skills. Cognitive skills are measured by the Raven score, and quintiles are of unequal size because the score is heavily tied. The sample includes $N = 13{,}407$ managers. Error bars show 95% confidence intervals.

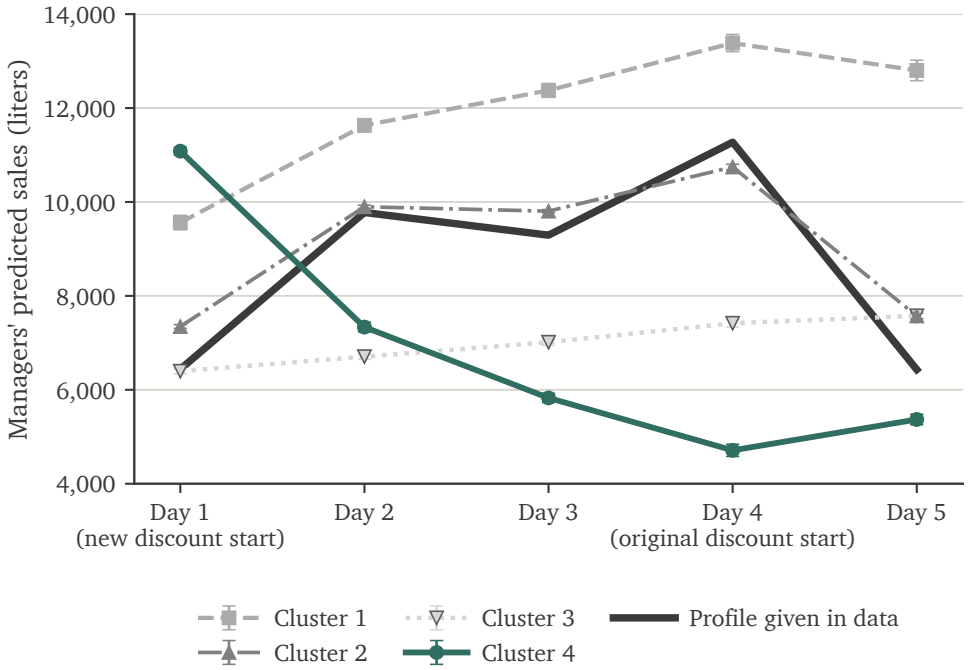
Additional results on the link between predictions in the detergent task and fuel price decisions

Figure E.7: Manager sales prediction clusters



Notes: This figure shows the result of k-means clustering with $k = 4$ on managers' sales predictions for the counterfactual pricing scenario. Each point is a manager, positioned by scores on the first two principal components, and the large outlined markers are the cluster centroids. Component 1 captures perceived elasticity of demand on days other than the discount start; Component 2 captures the degree of intertemporal substitution in the predictions, with higher values indicating more expected shifting of sales from Day 4, the original discount start, to Day 1, the new discount start. Figure E.8 shows the predicted sales profile of each cluster.

Figure E.8: Manager sales predictions by cluster



Notes: This figure shows managers' average predicted sales in the gas station counterfactual task, by cluster. Clusters are formed by k-means clustering with $k = 4$ on managers' five standardized sales predictions; Figure E.7 plots the clusters in the space of the first two principal components. The heavy line shows the profile of sales as presented in the original data. Error bars show 95% confidence intervals.

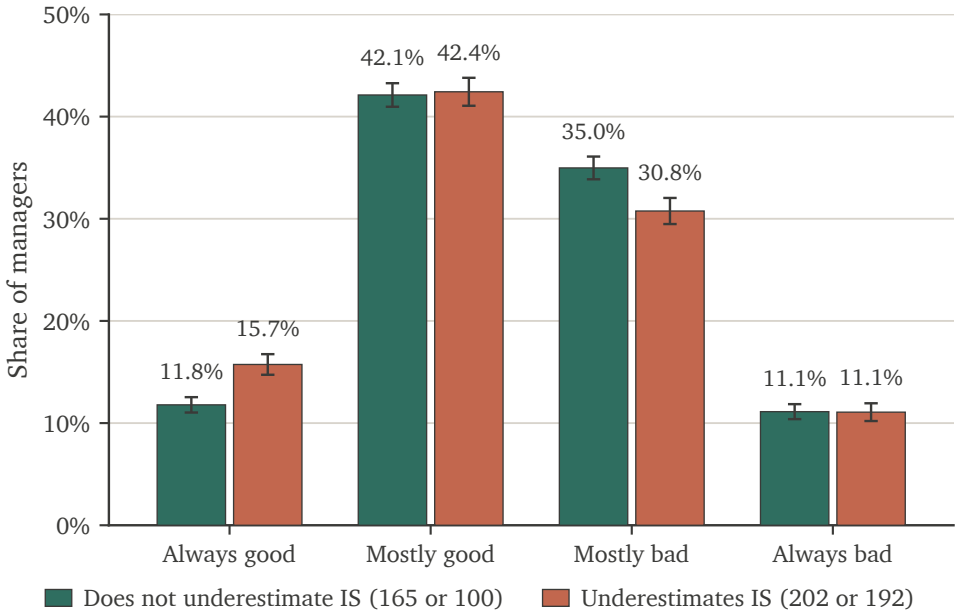
F Demand estimation and profit calculations

This appendix reports the demand estimates behind the profit calculations of Section 5.4 and the steps that turn the coefficients of Figure 12 into changes in daily fuel profit at an average station. The demand estimates, and the price scale used in Appendix F.2, are from Han et al. (2025a), which uses the same administrative data from our partner company.

F.1 Estimating the residual demand elasticity

We model each station as facing a downward-sloping residual demand curve that includes both consumers buying less fuel and consumers switching to competing stations, so that the elasticity incorporates competitors' responses in the local market. Following a standard approach in the gasoline demand literature (Li et al. 2014; Coglianese et al. 2017), we assume constant elasticity of demand and specify residual demand for station s as $q_s = A_s \cdot p_s^{-\varepsilon}$, where q_s is quantity sold, p_s is own price, $\varepsilon > 1$ is the price elasticity of demand and A_s is a station-specific demand shifter. Because we do not need to separate the direct response of demand from the cross-price response for this calculation, we adopt a single-instrument strategy that omits competitor price from the estimation and assume instead that the cross-price response is incorporated into ε.

Figure E.9: Mental models of intertemporal substitution and perceptions of profitability of price discounts



Notes: This figure shows the distribution of managers' beliefs about whether cuts of listed prices are good or bad for fuel profits, split by whether managers underestimate intertemporal substitution in the detergent task (predict 202 or 192) or not (predict 165 or 100). Error bars show 95% confidence intervals.

To estimate the elasticity, we take advantage of the fact that our partner company is vertically integrated. Under the company's internal accounting rules, the transfer price that determines each station's marginal cost moves one-for-one with the government-imposed price ceiling, which is indexed to world oil prices and adjusted every ten working days. We therefore use the ceiling as a cost shifter that is exogenous to local demand shocks. Let Ceiling_{jt} denote the price ceiling for fuel product j that is in force during the current pricing cycle t. Taking logs of the demand equation yields the estimating equation

$$\ln q_{sjt} = a_{sj} - \varepsilon \ln p_{sjt} + \delta_{sjm} + \phi_{sjw} + \psi_{sjy} + \eta_{sjt},$$

where q_{sjt} is the quantity sold in period t at station s of product j, p_{sjt} is the corresponding own price, a_{sj} is a station-product fixed effect, δ_{sjm} are month fixed effects, ϕ_{sjw} are day-of-week fixed effects, ψ_{sjy} are year fixed effects and η_{sjt} is an error term. We treat $\ln \text{Ceiling}_{jt}$ as the excluded instrument for the endogenous price regressor $\ln p_{sjt}$.

The price ceiling changes every ten working days, which creates variation at a frequency distinct from our time fixed effects. Day-of-week fixed effects absorb systematic differences across days within a week, month fixed effects capture seasonal patterns and year fixed effects control for annual trends. Because the ten-working-day adjustment cycle spans several days of the week and does not align with calendar month boundaries, we can include all of these fixed effects and still retain variation in the instrument that identifies the demand parameters.

We estimate an average elasticity of $\hat{\varepsilon} = 1.65$. As a check on the functional form, Appendix F.3 reports an alternative specification with linear demand, which separately identifies own-price and cross-price responses and yields an average elasticity of 1.63 in absolute value.

F.2 Profit calculations

Using the constant elasticity demand function of Appendix F.1, we calculate what the price gaps between groups of managers imply for daily fuel profit at an average station. We back out the demand shifter for an average station from the observed mean price and quantity, $\bar{A} = \bar{q} \cdot \bar{p}^{\hat{\varepsilon}}$, and compute quantities at counterfactual prices as $q = \bar{A} \cdot p^{-\hat{\varepsilon}}$, with $\hat{\varepsilon} = 1.65$.

Producer surplus is revenue minus variable cost,

$$PS = p \cdot q - MC \cdot q = (p - MC) \cdot q,$$

and changes in producer surplus equal changes in profit, because the difference between profit and producer surplus, namely fixed costs, is differenced out. Marginal cost is observed directly in the internal transfer price rather than inferred from prices. Let p_h and p_l denote the higher and the lower price, with q_h and q_l the corresponding quantities. The difference in producer surplus is

$$\Delta PS = PS_h - PS_l = (p_h - MC) \cdot q_h - (p_l - MC) \cdot q_l.$$

We set p_h equal to the average price $\bar{p}$ and p_l equal to p_h minus the estimated price gap, compute the corresponding quantity at each price, and report the percentage change in producer surplus as $\frac{\Delta PS}{PS_h}$.

The coefficients in Figure 12 are in standard deviations of the standardized monthly price ratio, which Section 3.1 defines, while the profit calculation needs a price gap in local currency per liter. In Han et al. (2025a), a coefficient of 0.017 standard deviations of the same price ratio corresponds to a price gap of 0.02 in local currency per liter at an average station, and we apply this scale to our coefficients. Relative to managers who predict 202, managers who predict 165 have a coefficient of 0.021 standard deviations ($p < 0.05$), a price gap of about 0.025 per liter, and managers who predict 100 have a coefficient of 0.025 standard deviations ($p < 0.05$), a price gap of about 0.029 per liter. The coefficient for managers who predict 192 is 0.014 standard deviations and is not statistically significant, so we report no price gap and no profit gap for that group. The price gaps are associations across managers conditional on the controls, not causal effects of holding a model, and the calculation compares managers who hold different models rather than changing the model of any one manager. With these price gaps, the lower prices of managers who predict 202

imply 3.3% lower daily fuel profit at an average station relative to managers who predict 165, and 3.9% lower daily fuel profit relative to managers who predict 100. The calculation is static.

F.3 Linear demand

The constant elasticity specification of Appendix F.1 does not separate the own-price response from the cross-price response. Linear demand separates the two and provides a check on the functional form. We use the conjectural variations framework for differentiated products (Slade 1986), in which residual demand for station i is approximated by

$$q_i = f_i(p_i, p_{-i}, z) = a + b_i p_i + c_i p_{-i} + g(z),$$

where q_i is the quantity of fuel products, gasoline and diesel, sold at station i, p_i is own price, p_{-i} is the average competitor price and $g(z)$ are demand shifters. To calculate how a price change affects producer surplus, we need the response of demand to price, $\frac{dq_i}{dp_i} = b_i + c_i\theta_i$, where θ_i is the competitor price response.

We use two exogenous cost shifters, the government-imposed price ceiling in the *current* pricing cycle, $Ceiling_t$, and the ceiling from the *previous* cycle, $Ceiling_{t-1}$, to estimate the own-price response b_i and the cross-price effect c_i. Two institutional features identify these responses. First, the price ceiling is updated every ten working days following a predetermined formula indexed to world oil prices. Second, our partner company's internal accounting rule mandates that marginal cost moves with the current ceiling *one-for-one*, which also means that the previous ceiling has little influence on the current price of our partner company's stations once the current ceiling is controlled for. Independent competitors purchase their refined oil from independent refineries, so their marginal costs adjust to the ceiling with a delay, which means that the marginal cost of competitors is affected by both the current ceiling and the ceiling in the previous cycle.

The estimating equation for station s and fuel product j is

$$q_{sjt} = a_{sj} + b\, p_{sjt} + c\, p_{-sjt} + \delta_{sjm} + \phi_{sjw} + \psi_{sjy} + \eta_{sjt},$$

where q_{sjt} is the quantity sold in period t at station s of product j, p_{sjt} and p_{-sjt} are the corresponding own and competitor prices, a_{sj} is a station-product fixed effect, δ_{sjm} are month fixed effects, ϕ_{sjw} are day-of-week fixed effects, ψ_{sjy} are year fixed effects and η_{sjt} is an error term.

The first-stage regressions show the expected differential impact of the two cost shifters. For our partner company's stations' own prices, the coefficient on $Ceiling_t$ is 0.98 (s.e. 0.001) and the coefficient on $Ceiling_{t-1}$ is 0.01 (0.001), consistent with near one-for-one

pass-through of current costs and a negligible influence of the lagged ceiling. For competitor prices the pattern differs: the coefficient on $Ceiling_t$ is 0.73 (0.001), indicating pass-through that is large but incomplete, and the coefficient on $Ceiling_{t-1}$ is 0.20 (0.001), reflecting the influence of previous-cycle costs.

The second-stage coefficients from the instrumental variables regression yield the own-price and cross-price elasticities evaluated at the mean price and quantity,

$$\bar{b}\,\frac{\bar{p}}{\bar{q}} = -2.07, \qquad \bar{c}\,\frac{\bar{p}}{\bar{q}} = 1.15.$$

An event study of competitor responses to price cuts that are not driven by a change in the price ceiling gives an average competitor price response of $\bar{\theta} = 0.38$. Combining the competitor price response with the demand parameters yields the effective price elasticity when a manager changes price while the price ceiling is fixed,

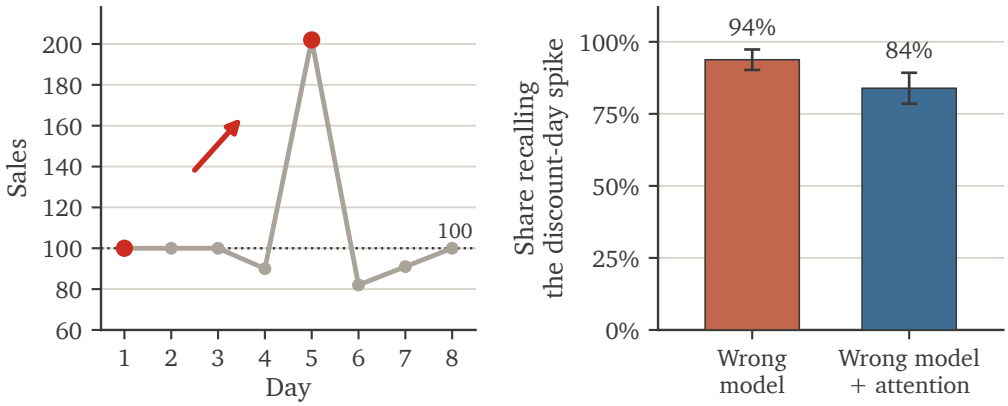
$$\frac{d\bar{q}}{d\bar{p}}\,\frac{\bar{p}}{\bar{q}} = (\bar{b} + \bar{c}\,\bar{\theta})\frac{\bar{p}}{\bar{q}} = -2.07 + 0.38 \times 1.15 = -1.63.$$

With this elasticity, the same price gaps of about 0.025 and about 0.029 in local currency per liter imply losses in daily fuel profit of 3.3% and 3.9%.

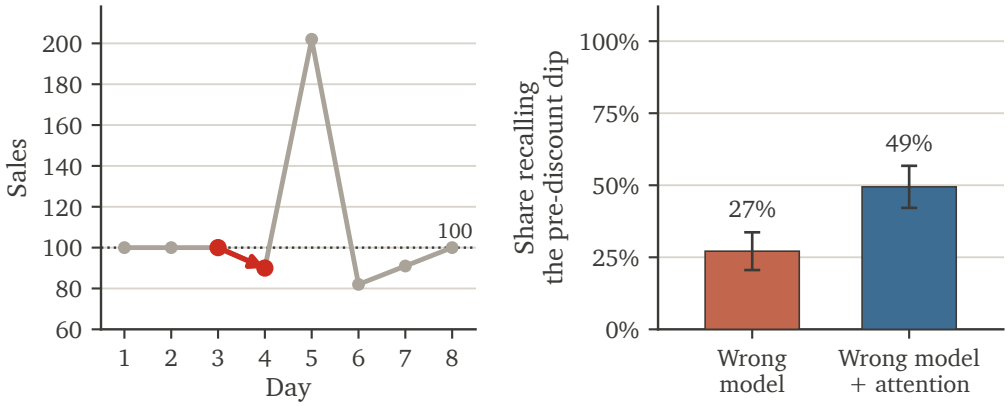
G Additional details and results from online experiments

Figure G.10: Recall of the three patterns in Study 1: *Wrong model* and *Wrong model + attention*

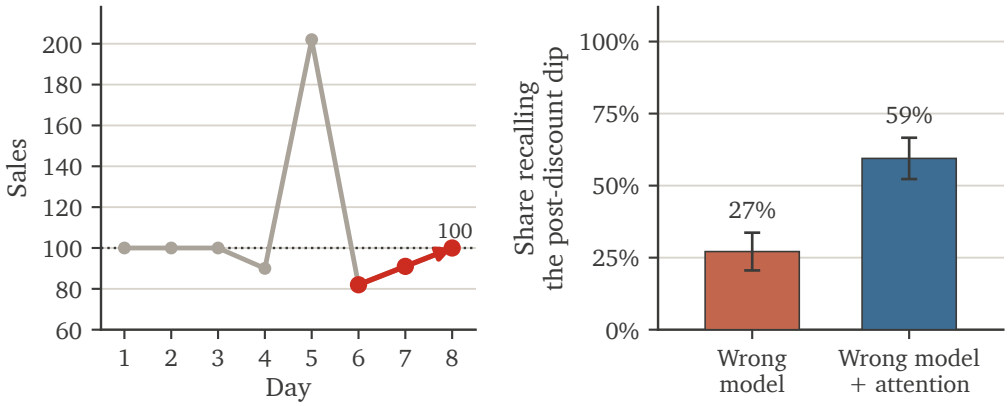
(a) The discount-day spike



(b) The pre-discount dip

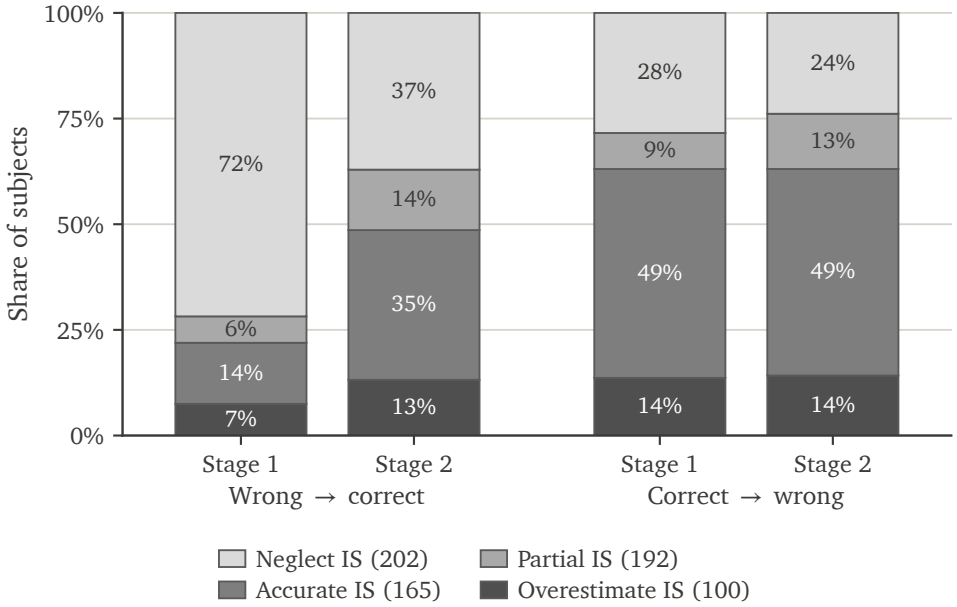


(c) The post-discount dip



Notes: This figure shows how often subjects recall each of the three patterns. Panel (a) shows the discount-day spike, Panel (b) the pre-discount dip and Panel (c) the post-discount dip. In each panel, the left plot shows sales on Days 1 to 8 and marks one pattern with an arrow, and the right plot shows the share of subjects who recall that pattern. The patterns are coded as in Figure 4. In *Wrong model + attention*, subjects answer two attention prompts about Days 3 to 4 and 9 to 10 and about Days 6 to 8 and 12 to 14 before the prediction. The sample includes $N = 177$ subjects in *Wrong model* and $N = 180$ in *Wrong model + attention*. Error bars show 95% confidence intervals.

Figure G.11: Distribution of predictions in each stage of Study 2



Notes: This figure shows the distribution of best guesses of daily sales under a permanent price of 5.2, by group and stage. The correct answer is 165, and 202 neglects intertemporal substitution. Wrong → correct: subjects assigned the wrong model first and the correct model second ($N = 720$). Correct → wrong: subjects assigned the correct model first and the wrong model second ($N = 176$).

Samples and balance. Table G.1 reports the analysis samples and subject characteristics in each treatment. The exclusions are similar across treatments. In Study 1, 17, 11 and 13 subjects in *Wrong model*, *Wrong model + attention* and *Correct model* ticked "I don't recall" for all seven days, and 4, 6 and 0 were flagged as using an AI tool. In Study 2, the corresponding numbers are 29, 35 and 21, and 11, 9 and 2, for the treatment without the data shown again, the treatment with the data shown again and the treatment that assigns the correct model first. In Study 1, subjects in *Wrong model* are older on average and less likely to be women than subjects in the other two treatments. Controlling for age, gender and Raven score, the effect of the wrong model on the share predicting 202 is 43 percentage points and the effect of the attention prompts is −13 percentage points (both $p < 0.01$), compared to 42 and −11 percentage points without controls. Among all 600 subjects who completed Study 1, the two effects are 38 and −9 percentage points ($p < 0.01$ and $p < 0.05$). Among all 1,003 subjects who completed Study 2, the difference between the changes in the two orders is 31 percentage points ($p < 0.01$), and showing the data again changes the share predicting 202 in stage 2 by less than one percentage point.

Predictions typed into an empty box. In Study 1, subjects choose their prediction from four answers, and the menu could suggest a model to a subject who would not otherwise have one. We therefore ran a follow-up pilot of Study 1 in August 2026 in which subjects type the prediction into an empty box and earn $1.00 if the value is within 10 of 165. The pilot has two treatments, *Wrong model* with 115 subjects and *Correct model* with 114. The pages,

Table G.1: Sample size and subject characteristics by treatment

	N	Age	Female	Raven score
Study 1				
Wrong model	177	40.8	0.46	4.65
Wrong model + attention	180	37.6	0.60	4.48
Correct model	192	38.6	0.59	4.55
p-value, equal means		0.05	0.02	0.76
Study 2				
Wrong → correct, no data in stage 2	367	39.5	0.57	4.50
Wrong → correct, data shown again	353	39.5	0.58	4.57
Correct → wrong	176	39.1	0.60	4.62
p-value, equal means		0.95	0.89	0.80

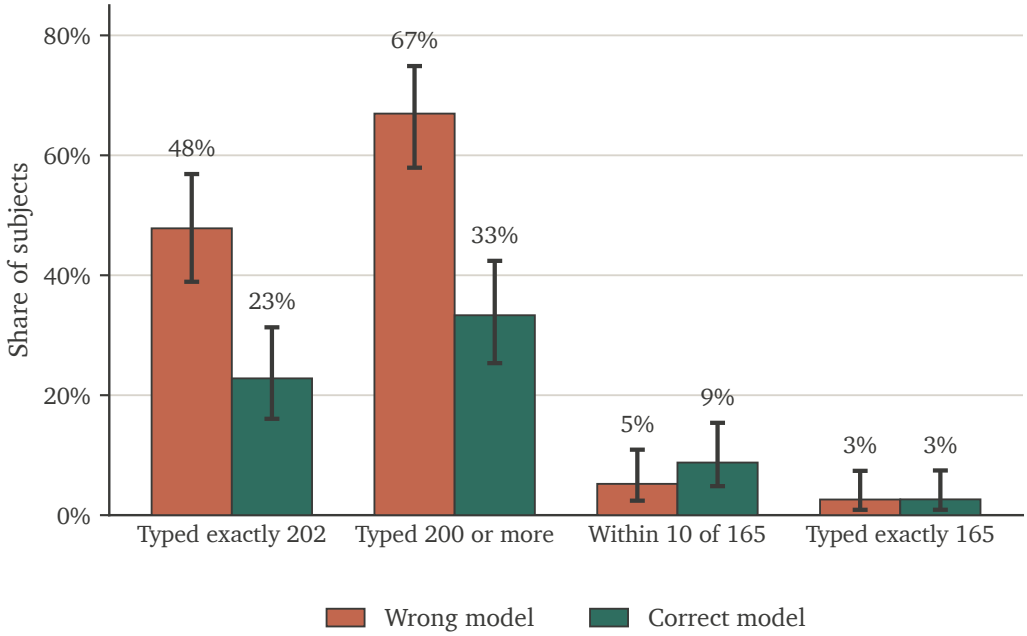
Notes: This table reports the analysis samples after exclusions. Age and gender are taken from subjects' Prolific profiles, and Female is the share reporting female. Raven score is the number of correct answers to nine Raven matrices. The p-values are from F-tests of equal means across the treatments of each study.

the descriptions, the recall task and the Raven matrices are those of Study 1, and a question about confidence in the typed prediction, which is not incentivized, replaces the allocation of 100 points. Figure G.12 shows the predictions. The wrong model raises the share of subjects typing exactly 202 from 23 to 48 percent, and the share typing 200 or more from 33 to 67 percent (both $p < 0.01$). About 12 percent of subjects in *Wrong model* type 200 or 203, values the menu does not offer. The wrong model also lowers recall of the pre-discount dip by 13 percentage points and recall of the post-discount dip by 16 percentage points (both $p < 0.05$), while recall of the discount-day spike and the placebo differ between the treatments by less than 4 percentage points. Few subjects in either treatment type a value close to 165: 5 percent in *Wrong model* and 9 percent in *Correct model* are within 10 of it, and the difference is not significant.

G.1 Attention measured by clicking

Recall measures attention after the data are gone. An earlier pilot measured attention while subjects worked with the data. Subjects saw the same table of 14 days, with every sales value hidden behind a button. Clicking a button revealed that day's sales for one second, after which the value was hidden again, so a subject who wanted to hold two days side by side had to click repeatedly. We count the clicks on a day as attention to that day. The dip days are the days of the pre-discount and post-discount dips, Days 4, 6, 7, 10, 12 and 13, and the baseline days are Days 1, 2, 3, 8, 9 and 14, on which sales are 100. Subjects were randomly assigned the wrong model or the correct model, as in Study 1, and the pilot had no attention prompts. We use the treatments in which clicking is free, and leave out a further treatment that charged \$0.05 per click.

Figure G.12: Predictions typed into an empty box

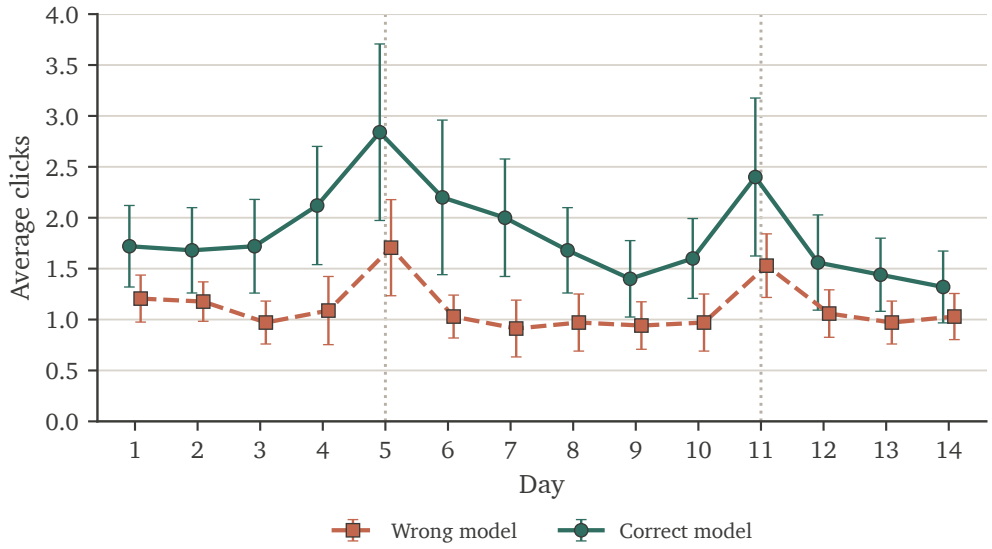


Notes: This figure reports results from a follow-up pilot of Study 1 in which subjects type their prediction of daily sales under a permanent price of 5.2 into an empty box. Each pair of bars gives the share of subjects whose typed prediction meets the condition below it. Subjects earn $1.00 if the typed value is within 10 of 165. Error bars show 95% confidence intervals. *Wrong model* $N = 115$, *Correct model* $N = 114$.

Figure G.13 plots the average number of clicks on each day. Subjects assigned the wrong model click the dip days less than subjects assigned the correct model, 1.01 against 1.82 clicks per day ($p < 0.01$). These subjects also click less on every other kind of day, 1.62 against 2.62 on the discount days and 1.05 against 1.59 on the baseline days (both $p < 0.05$), and they click 15.6 times in total against 25.7 ($p < 0.01$), so part of the difference is in how much of the table subjects inspect at all. The contrast between the dip days and the baseline days separates the two. Subjects assigned the correct model click the dip days 0.23 times per day more than the baseline days, while subjects assigned the wrong model click them 0.04 times per day less, and the difference between the two contrasts is 0.28 clicks per day ($p < 0.05$). The share of a subject's clicks that falls on the dip days is 0.36 under the wrong model and 0.39 under the correct model, a difference that is not significant.

The pilot is small, and clicking imposes a cost of attention that the recall measure does not. We report it because it measures attention while the data are on the screen rather than afterwards, and because it points the same way as the recall results in Section 4.1.2.

Figure G.13: Clicks on each day by assigned model



Notes: This figure shows the average number of clicks on each day's sales value, by assigned model, in the pilot treatments in which every value is hidden and a click reveals one value for one second. The dip days are Days 4, 6, 7, 10, 12 and 13, and the dotted lines mark Days 5 and 11, the discount days. Error bars show 95% confidence intervals. The sample includes $N = 34$ subjects assigned the wrong model and $N = 25$ assigned the correct model, and leaves out the treatment that charged for clicks.