

参赛论文封面

参赛学生姓名： 徐靖尧

中学： 上海市世外中学

省份： 上海

国家/地区： 中国

指导老师姓名： 吕劭炜

指导老师单位： 上海市世外中学

论文题目： Conversion or Discovery? Consumer Search

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Conversion or Discovery?

Consumer Search and Ranking Design in Online Food Delivery

Jingyao Xu

Abstract

Online platforms select which product a consumer sees first so they can increase the chance she buys something. This paper asks if that objective serves consumers and shows it does not. We analyze one million food delivery orders of 200,000 consumers in Beijing by observing what each consumer ordered and which restaurants she browsed first.

We started from the assumption that consumers know very little about most restaurants other than a few standouts. The two ways a consumer resolves the same uncertainty are browsing and prior experience—whichever is cheaper. This has a surprising effect, which the data confirm: A restaurant a consumer finds while browsing is less likely to be a regular choice than one she orders from while not browsing. Comparing the same consumer's own orders shows a difference of 13 percentage points. Not that browsing is useless. Browsing takes place precisely where the consumer knows least, so what it finds is more uncertain, and more uncertain choices fail more often.

We first test alternative search protocols and find that the observed behavior is inconsistent with canonical sequential search but consistent with fixed-sample-size search. On that protocol we estimate its search-cost parameters from the observed browsing distribution and compare different restaurant display rules. Showing a consumer a restaurant has two parts to its welfare value: the chance she orders from it and the worth of finding out whether it suits her. A ranking built on purchase probability captures the first precisely, and the second not at all. By reordering the display to take into account both, expected match quality is increased by 6.9 percent at our baseline calibration and by 2 to 21 percent over the range we take into account. Those that such a ranking under-shows are the restaurants a consumer knows least about and that are about as good as what she orders, exactly the ones worth finding out about.

Keywords: consumer search, platform design, ranking algorithms, experience goods, food delivery

JEL codes: D83, L81, L86, D12

1 Introduction

There are thousands of restaurants to choose from when you open a food delivery app, and you’ve probably got a minute or two to make your choice. You almost always have a short list of places that you’re familiar with, or want to continue having relationships with, and almost always you don’t know anything else about restaurants other than their star rating, photos, and price. Whether you get curious enough to look into one of the unfamiliar restaurant options depends, at least in part, on how much time it takes to do it and how much use you expect to get out of it.

In making such a decision, the food-delivery platform is anything but a passive observer. The platform’s ranking algorithm determines which of the restaurants appears first, and which of them the consumer sees at all. Such recommenders are being trained to maximize the probability of “conversion” – the chance the consumer places an order. This question is of course commercially very natural, but in this paper we ask whether this objective is optimal – we show that it is not, in a specific and measurable sense we characterize theoretically and quantify empirically.

Consumers do, however, differ in their prior knowledge about products. Traditional models of consumer search (Stigler, 1961; Weitzman, 1979) usually have in common an implicit assumption that the prior knowledge of all unsampled alternatives in the product market is identical at the outset. As we discuss later, this implicit assumption is badly violated in the area of consumer search considered here, the repeat-purchase market for experience goods. A consumer who orders from a restaurant four times will know very precisely her own match with that restaurant after those four orders. The unsampled restaurant from which she has not yet ordered has, from her point of view, no known fit at all, and its only characteristic is that it has some average restaurant rating, such that its unobserved fit for the present ordering decision is only poorly correlated with its average rating. To put it another way, the prior precision of alternatives in the consumer’s set of possible alternatives is neither common to nor independent of the alternatives in the consumer’s set of possibilities.

Because search and self-selection are substitute technologies for reducing uncertainty, there are important consequences to correctly using search in the presence of preexisting heterogeneity in how well consumers understand their own preferences. This is the direction of influence most people, when they think about how to design a platform, do not anticipate. We focus on one such example: search for matches to alternatives about which the consumer has less prior knowledge. The systematic consequence of deploying search of this sort is that search generates matches that fail with higher probability. In general, an unbiased Bayesian consumer who must later decide whether to return the match (without knowing precisely why in advance) finds those alternatives more dispersed, whereas dispersion of alternatives imposes costs.

What we do

We address these questions using the Takeout Recommendation Dataset: a public release of 1 million order-level records placed by 200K identified consumers across 11 commercial districts of Beijing over three weeks (March, 2021) in the largest food delivery platform in China, Meituan. Importantly, our data provide the restaurant browsing sequence of a consumer before they make an order, which is also made public. Combined with ordering history, we are able to jointly model search and choice.

The paper has three parts.

A model with heterogeneous prior precision. We consider a fixed-sample-size search problem in which the variance of the consumer’s prior over match value σ_{ij}^2 is not only indexed by the alternative j but also by the consumer i , and is decreasing in prior consumption x_i . Inspection is informative, though not decisive; like restaurants, this is an experience good. Two additional predictions follow from the model: Intensity of search, on the margin, is increasing in prior uncertainty, so private information and search are substitutes (Proposition 1). Given that a particular choice has been made the probability that it will be a repeat purchase is decreasing in the residual uncertainty after inspection (Proposition 2). Moreover, the value a platform derives from displaying an alternative can be understood as being made up of a conversion-value component and an option-value component; a ranking criterion that maximises the former component alone is what we call a conversion criterion (Proposition 3).

Testing the search protocol. Before the choice of one model class over another, we need to figure out which one of them fits our data. In this section, we provide a full test battery of protocols like the ones tested by De los Santos et al. (2012), who studied online book purchases in a quite different market. Our data reject sequential search in favour of fixed-sample-size search on all the tests proposed there. Purchased restaurants represent the last restaurant of the session only 38.4% of the time (against a random benchmark of 23.2%) and for genuine recall in 39.1% of cases. Also, and in a much stronger way, the decision to keep browsing is not correlated with the best restaurant encountered so far when conditioning on the number of options that have already been inspected. As a final test, that does not have an equivalent test in the original article, we report that the order in which restaurants are inspected by users is essentially uncorrelated with their quality. We then interpret this fact not as an example of consumer irrationality but as a measurement of the objective of the platform.

The durability of brand knowledge and the search process. Our central empirical finding is that restaurants uncovered without resort to search are a great deal more likely to be purchased by consumers repeatedly than those found using search. The difference among consumers is 13.3 points ($t = -54.2$) and it is monotone with the intensity of search. The effect survives the inclusion of restaurant fixed effects with minimal loss, and is unaccompanied by any change in the observed quality of restaurants found through the search process. Consumers are not driven toward worse restaurants through search. Instead, the restaurant found through search should have lower durability as compared

with the non-search restaurant, just as predicted by Proposition 2. This is the reason the observed correlation between search intensity and match quality is not, in itself, a measure of the return to search.

Why the ranking result matters beyond food delivery

Proposition 3 yields a partitioning of optimal choice that we think is the paper’s most portable contribution. Denote by a_j the standardised gap between the consumer’s incumbent value v and alternative j , $a_j = (v - \delta_j)/\sigma_j$. The net value of displaying j then decomposes as

$$MV(j) = \underbrace{\sigma_j \varphi(a_j)}_{\text{option value}} + \underbrace{(\delta_j - v)[1 - \Phi(a_j)]}_{\text{conversion value}}.$$

Thus, the probability that any specific alternative j will convert is exactly $1 - \Phi(a_j)$. So a platform that ranks pages by their predicted purchase probability is ranking them by $\partial MV/\partial \delta_j$ while ignoring $\sigma_j \varphi(a_j)$ altogether. The omission is largest when $\delta_j = v$, that is when $a_j = 0$. The alternatives overlooked are just those that are both unfamiliar and similar to what the consumer orders. These are exactly the ones discovery is most likely to uncover.

But this argument isn’t specific to restaurants. It applies anywhere a platform tries to rank alternatives (differentiated goods and services) for consumers who don’t know beforehand which option will satisfy them best – hotels, architects, streaming services, app marketplaces, and more.

Roadmap

Section 2 places the paper in the literature. Section 3 describes the setting and data. Section 4 presents descriptive patterns that motivate the model. Section 6 develops the theory, with proofs collected in Appendix A. Section 5 tests which search protocol the data support. Section 7 presents the durability results, Section 8 the heterogeneity across consumers, and Section 9 robustness. Section 10 draws out the implications for ranking design. Section 13 states what the paper cannot establish, and Section 14 concludes.

2 Related Literature

This paper relates to three literatures: an empirical investigation of consumer search, economics of platform design, and the literature on consumer inertia in markets with repeated purchases.

Empirical consumer search. After Stigler (1961) it was established that costly information acquisition (the search itself) breaks any equivalence between nominal competitiveness and actual competitive outcomes. Empirical studies of consumers then tended to use observed price dispersion, for example Sorensen (2000) for prescription drugs or

Brown and Goolsbee (2002) for term life insurance, to infer underlying consumer search frictions. However, more recently researchers have also started looking at demand to infer consumer search and even to observe search. A prominent example is the recent work by De los Santos et al. (2012), where the authors consider the question of whether consumers do sequential searching or use a predetermined size search sample as they visit web pages and purchase books. They reject sequential search. Using the same test (but in another market), we do the same here. Our departure from much of the relevant existing literature is that rather than assuming that the unobserved component has a fixed degree of precision, we model the degree of prior precision that consumers have for alternatives that they have not experienced and accumulate it as a result of consuming different products.

Design of platform. Dinerstein et al. (2018) provide evidence that platform design serves two potentially unrelated purposes, on the one hand to aid consumers to find products they prefer, and on the other to enhance competition between sellers. Our ranking result is along similar lines but considers a different wedge. We suppose the seller side is fixed and ask how an objective ranking for the purchase probability affects the composition of what consumers look at. Bergemann and Bonatti (2024) discuss in a theoretical context how a platform’s choices around data and disclosure influence market matching and welfare. Our contribution is to provide a decomposition that can be empirically identified using the observed distribution of searches.

Consumer inertia. Dubé et al. (2010) find strong evidence that repeat purchasers of packaged goods do so because they are loyal rather than being lazy, slow to learn or engaging in some other kind of search. Their argument is only partly persuasive, because their data track only actual purchases. Because we observe browsing and, hence, a search process, we see the same pattern there as here: more browsing is strongly associated with less repeat purchase. We do not claim to contradict their conclusion, since in a cross section no causal relationship can be established between searching and brand switching. What we claim is that this important “unobservable” is actually quite large in a high frequency market, and that models in which search and private information are substitutes organize the data much better than models in which inertial preference for one’s prior choices acts as a preference coefficient.

Experience goods. As a class, restaurant foods fall into the category of “experience goods,” meaning that the most important characteristics, such as taste, amount, and packaging, cannot be evaluated prior to actual consumption. Jin and Leslie (2003) show that requiring mandatory quality disclosures changes the way consumers sort themselves and forces restaurants to change the products they offer. Our setting allows us to separately report the two dimensions of quality, food quality and delivery quality, which we find are only weakly correlated in the data. Because of that, it is an empirical question

(which we note in Section 10 but do not resolve) which dimension to prioritize and surface as the overall platform ranking.

3 Setting and Data

3.1 Institutional setting

Meituan is China’s biggest food delivery platform and an online consumer can use it to discover, book, and order food. When a consumer opens an application, it shows her a ranked list of restaurants that can deliver to the consumer’s address. A consumer can also browse through the restaurant’s page. After that, Meituan will assign a courier to carry the consumer’s order from the restaurant.

The setting has two aspects that matter for the analysis. One is the high rate of purchase frequency. In our sample, consumers have a median frequency of four per three week period. Thus, consumers are rapidly accumulating private information about the restaurants in question so that it is reasonable to assume this is the dominant source of variation in prior precision. Another is that consumption is constrained in time, and that consumption is constrained in a systematic way within the day. Ordering lunch at 11:55 on a workday imposes a stronger time constraint on the amount of time available for browsing on a particular day than does the action of ordering dinner at the same restaurant on a Saturday night for a particular day. This variation in the time constraint and therefore search cost is primarily used for shifting behavior.

3.2 Data

We use a publicly released Chinese takeaway order data set, the Takeout Recommendation Dataset (TRD).¹ It records eleven commercial areas in Beijing between 1 and 28 March 2021, with the first three weeks forming our main data to be used throughout this paper. As we only use data up to 21 March 2021, the remaining week of data in the original release is un-used.

Based on a sample of 1,068,495 orders from 200,000 customers at 3,252 restaurants, we describe here key facts about the order process. Since everyone placed at least two orders, the sample is properly a short panel, and not a cross-section. Descriptive statistics appear in Table 1.

In all cases, we either limit the reported value to a preassigned ordinal band or specify the type of information that is released. The release of actual continuous item prices is restricted: we report the item-level price values from the food catalogue where the continuous price measure is required, but where a discrete categorical measure is appropriate

¹Data collected for this research is deposited in Zenodo (doi: 10.5281/zenodo.8025855) under a Creative Commons Attribution 4.0 licence. This data set was compiled for the purpose of analyzing recommender systems. To the best of our knowledge, this is the first application of this dataset in economic context.

Table 1: Summary statistics

	Mean	Median	p90	N
<i>Panel A. Orders</i>				
Orders per consumer	5.34	4	11	200,000
Distinct restaurants per consumer	3.75	3	7	200,000
Food items per order	3.22	3	6	1,068,464
Repeat order (=1 if consumer ordered here before)	0.298	—	—	1,068,495
<i>Panel B. Search</i>				
Distinct restaurants inspected	2.58	0	7	1,299,045
Any inspection recorded (=1)	0.472	—	—	1,299,045
<i>Panel C. Restaurants</i>				
Overall rating	3.85	—	—	27,076
Food quality rating	3.40	—	—	27,076
Delivery quality rating	4.03	—	—	27,076
Item price (RMB)	28.40	20.00	57.50	179,759

Notes. Panel A depicts the training sample over a three-week period; browsing records cover all 28 days. The restaurant and item catalogues for Panel C are from restaurants that were rated at the time of our observation – restaurants rated zero were instead treated as missing observations. The correlation between food and delivery ratings of restaurants that we have ratings for is only 0.306.

we only report the ordinal consumer categories based on past spending. We are particularly careful here not to interpret these ordinal spending bands as income, and explicitly state that our measure is derived directly from past spending.

3.3 Measurement of search

For each order, we recorded the sequence of restaurant pages that a consumer opened before placing the order. To capture “intensity” of search, we counted the number of distinct restaurants on those pages. (We count a consumer who opens the same page 4 times as making one, not four, inspections.)

3.4 A limitation that shapes the design

We explicitly note that our browsing log is far from exhaustive. In those browsing sessions where the customer visited at least two restaurants, the restaurant the end consumer actually ordered from appeared as part of the recorded browsing sequence only 19.4% of the time. This suggests that consumers reach at least a significant minority of restaurants via channels unlogged here, such as text search, favouritising certain restaurants, and knowing in advance where they intend to order from.

So we take three steps. First, we point out that in our study, the unit of analysis is a customer who logs her preference for a firm—not a representative consumer, whose preferences may include choices she does not make online. Second, we note that a key feature of our search protocol (see Section 5), the conditional on observed inspection, ensures that we can only learn about behavior that we actually observe. Thus, the most

we can conclude is that a significant fraction of our results apply to unobserved behavior; at the very least, our estimates cannot be overstated. To address this concern, we re-estimate all our results separately within each decile of search completeness; the central result remains robust, albeit weaker, where observation rates are the strongest.

4 Descriptive Facts

4.1 Fact 1: demand is sharply time-structured

Order volume for the entire day is strongly bimodal, peaking at 11:00 with a high of 208,763 orders and again at 18:00 with about 108,425, against troughs of roughly a thousand per hour between 03:00 and 05:00. These meal-period peaks are also picked up by the platform’s own classification of meal periods.

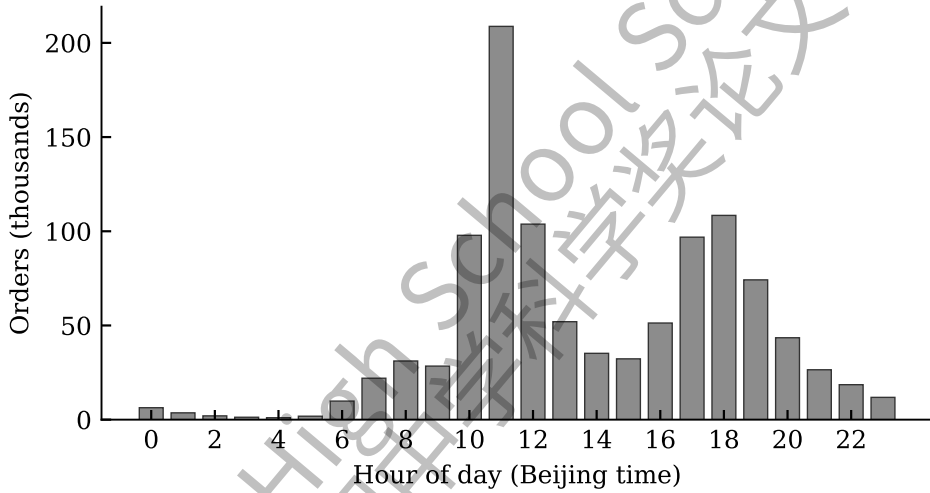


Figure 1: Order volume by hour of day. Beijing time; three-week training sample.

Order values and the proportion of orders greater than 65 RMB increase monotonically throughout the day: 6.5% at breakfast, 16.3% at lunch, 20.8% in the afternoon, 24.8% at dinnertime, and 19.0% late at night. The circumstances of the meal occasion therefore differ, so both deliberation time and stakes in decision-making vary.

4.2 Fact 2: markets are unconcentrated but demand is local

The degree of restaurant concentration is quite low. Herfindahl indices in commercial districts range from a low of 18 to a high of 54 out of a possible 10,000, and the ten largest restaurants in a district range in their percentage share of total orders from 6.2% to 15.8%. By any measure these are atomistic markets.

At the same time, orders are very local in that only 5.0% of orders came from outside a consumer’s “modal delivery district”. For example, 77.8% of consumers specified only one delivery address for the entire sample period. The eleven districts thus constitute largely separate markets, and we treat them as such.

This juxtaposition is worth commenting on. Although aggregate industry concentrations are very low, the median consumer orders from just three different restaurants in a three-week period. Whatever limits effective competition in this instance is operating at the level of individual customer consumption, rather than in the market structure.

4.3 Fact 3: inertia accumulates

At each order, 29.8% are to restaurants that a consumer has previously placed an order with. This percentage climbs quickly as a consumer orders more: 22.2% at his or her second order, 34.1% at the fourth, 43.4% at the seventh and 49.3% at the tenth. Inertia is not something a consumer brings with him or her, but rather something that develops over time.

The degree of repeat-buying also varies from one eating occasion to another. Again, this cannot be reasonably captured within the model of repeat-buying based entirely on taste-preference. The repeat-buy rates peak at breakfast (33.8%), trough late night (27.2%) and vary between lunch, afternoon and dinner. The same consumer buys more habitually at eight in the morning compared to nine in the evening.

4.4 Fact 4: Increased expenditure does not mean increased searching

The fourth pattern is the one that spawned this paper. Once customers are divided into historical “bins” based on cumulative expenditures, the repurchase rate declines monotonically from 34.7% for the low band to 25.5% for the top band. Higher spenders have a higher exploration ratio.

Search intensity is about the same for all groups in the bands: the mean number of distinct page inspections (that is, page views) is, in increasing price-band order, 2.76, 2.76, 2.74, 2.64 and 2.51. In other words, higher-spending consumers search no more than others, and if anything slightly less.

Table 2 explains many consumer habits. For consumers who never search in the first place, the percentage exploring increases from 60.0% to 71.8% as expenditure increases. But the “extra” exploring that purchase of search buys – which amounts to only 6.0 percentage points for the highest-spending consumers in the sample – is twice as valuable to the lowest-spending consumers as to the highest.

However, a simple search-based model of exploration does not fit these facts jointly. If, indeed, search is a route to exploration, then exploration and search should be closely associated. But they are not. The most natural reading of these data is that search is just one way of becoming willing to try something unfamiliar, and prior knowledge is an alternate route; those who have had a little of the second will require much less of the first.

Table 2: Exploration and search by consumer spending band

Spending band (RMB)	<29	29–36	36–49	49–65	≥65
Mean realised order value	26.73	32.74	39.65	48.40	59.28
Repeat-order rate	0.347	0.312	0.293	0.274	0.255
Mean distinct inspections	2.76	2.76	2.74	2.64	2.51
Pr(explore), did not search	0.600	0.636	0.664	0.688	0.718
Pr(explore), searched	0.705	0.742	0.752	0.767	0.777
Search premium	0.106	0.106	0.088	0.078	0.060

Notes. The platform’s spending bands classify consumers by past spending; they are not income. That they are economically meaningful is confirmed by the realised order value per customer, which increases monotonically from the lowest band to the highest (first row). Search behaviour is given by the exploration rate with and without any recorded inspection, and the search premium is the difference between them (last three rows). $N = 1,068,495$.

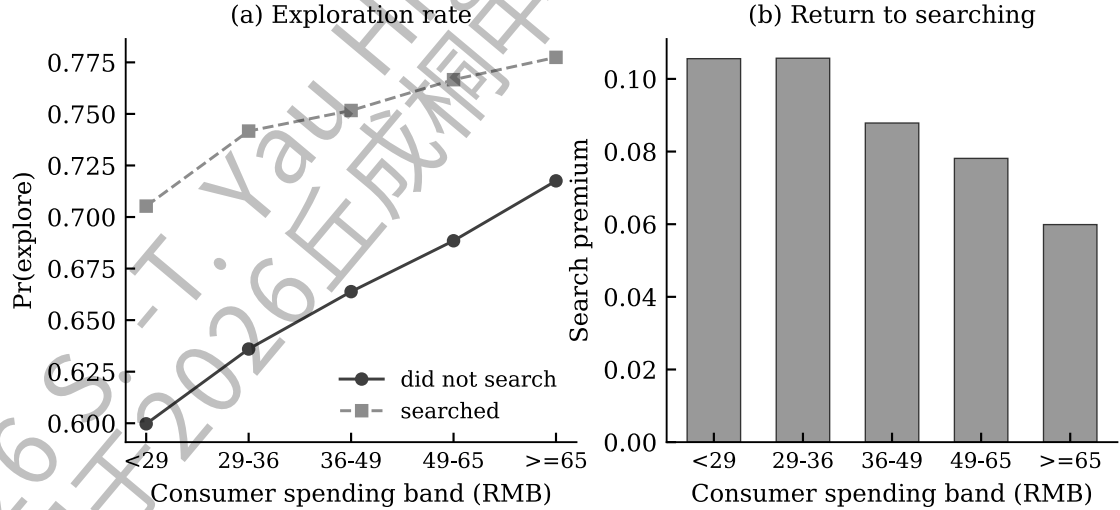


Figure 2: How and how much they explore (and return for more). Panel (a) plots the probability that the order is placed at a restaurant not previously tried, for orders with and without inspection. Panel (b) plots that difference.

5 Which Search Protocol Do the Data Support?

To develop a search model that is consistent with the data, we first show that the class of sequential search models in the literature is inconsistent with browsing behavior. De los Santos et al. (2012) show that browsing data can distinguish sequential search, in which the consumer inspects one option at a time, and decides after each whether to continue, from fixed-sample-size search, in which the consumer commits to a sample size (or number of inspections) in advance. The two protocols have different observable implications, and we test three of them.

These tests estimate average search on the food delivery platform’s recommendation surface. For example, the estimation sample for the recall test includes sessions where at least two different restaurants were inspected and purchased. The recall estimation sample includes 83,381 orders; the ordering test uses 306,450 sessions that had three or more restaurant inspections. As we explain in Section 3, the samples are selected and the tests should be read as describing search on the platform’s recommendation surface.

5.1 Test 1: recall

Under this sequential search procedure, the consumer will choose as his final purchase whatever he encounters that seems good enough at the end of his search sequence; in other words, his final purchase should always be the last option he inspects. Under the fixed sample size search procedure, on the other hand, the consumer inspects a set amount of options and, regardless of their ranks, simply picks the one that seems best. Consequently, his final purchase may occur anywhere in the set he inspects.

In about 38% of cases, the restaurant you purchased is the last one inspected. (A random benchmark for that proportion is $1/n$, where n is the number of inspections you have had; here it is 23.2%.) That is well above chance, but far below the unit share a sequential stopping rule implies. Genuine recall, i.e., purchasing an option that is neither the first nor the last inspected, happens in 39.1%.

Table 3: Position of the purchased restaurant in the inspection sequence

Distinct inspections n	2	3	4	5	6	7	≥ 8	All
Share purchased last	0.602	0.481	0.421	0.384	0.348	0.316	0.237	0.384
Random benchmark $1/n$	0.500	0.333	0.250	0.200	0.167	0.143	—	0.232
Observations	14,643	12,242	9,731	7,434	6,144	4,963	28,224	83,381

Notes. Sample: A plurality of purchases occurred at restaurants that were neither the first nor the last option inspected during the session. Of the 83,381 sessions, 18,695 (22.4%) were the first restaurant inspected, 32,054 (38.4%) the last, and 32,632 (39.1%) neither first nor last.

Table 3 shows that the purchase pattern with respect to the length of the inspection sequence deteriorates steadily. At two inspections it is still more than half the time (60.2%). Also notable is the high rate in long sessions (23.7% at eight or more inspections, as opposed to roughly 12% under the random benchmark). However, the increase for

recently inspected restaurants seems small and therefore hard to reconcile with a protocol where stopping = purchasing.

5.2 Test 2: does continuation depend on what has been found?

That first test has a special content for sequential and fixed-sample size searches. For sequential search, in which the decision to continue the search is made after each observation, a natural reservation-value rule dictates continuation if any uninspected alternative is higher. However, for a fixed-sample-size search, the decision to stop must be made without knowing anything about what is to be found, so the number of samples n is chosen prior to inspection.

For each session and each position k in that session, we create a flag for continuation after k and the best observed rating among the first k inspections. Ordering observations by the size of the best so-far, the probability of continuing decreases monotonically from 0.947 in the lowest octile to 0.805 in the highest octile. In other words, taking the data at its face value is evidence in favor of sequential search.

If we plot best quality so far against the number of inspections of the same session, there is a positive correlation, as longer sessions mechanically will have higher best-quality-so-far. However, once we control for inspection position (k) then the correlation should be zero as should be any monotonic relationship. Conditional upon $k = 3$, the probability to continue in all quality quartiles is, for example, 0.770, 0.764, 0.757 and 0.771. There is no meaningful monotonicity. When conditioning on $k = 4$, the probability is 0.786, 0.779, 0.774 and 0.780. Hence any apparent correlation is an artifact of data construction.

The contrast of panels (a) and (b) of Figure 3 provides important evidence. If one were to look only at panel (a), he or she might conclude that consumers employ the reservation-value rule described above. Panel (b) demonstrates otherwise.

5.3 Test 3: is inspection ordered by attractiveness?

The index policy of Weitzman (1979) dictates searching through alternatives in order of declining reservation value. Reservation value should increase monotonically with both the expected quality of the restaurant and the uncertainty about its quality. Hence, if consumers chose the inspection order, we should see a negative correlation between the position at which one inspects a restaurant and its quality.

So it hardly does it. Across 306,450 sessions with three to 20 inspections, the mean within-session correlation is -0.0085 ($t = -8.9$), with a median of -0.0092 ; only 50.6% of sessions produced a negative correlation, compared with exactly 50% for pure randomness of outcomes. The estimate is statistically distinguishable from zero, but economically negligible.

Unlike the previous two tests, we look at this differently in a two-sided market where the consumers don't have to walk down a street to make an informal inspection of many

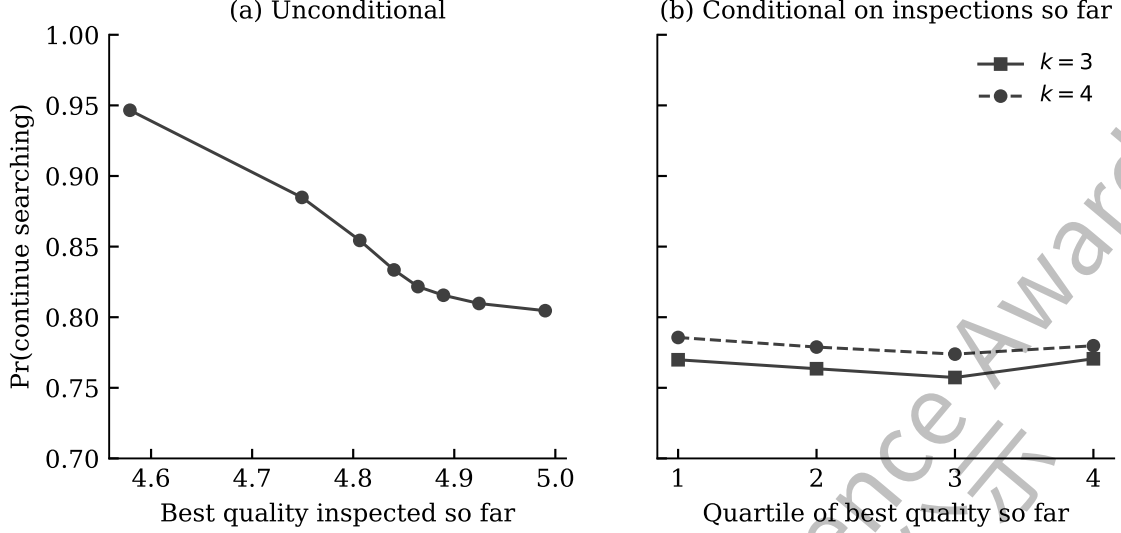


Figure 3: The stopping-rule test: Panel (a) shows the probability that a diner would continue to search given the restaurant rating received to that point (averaged across all search positions). In panel (b), the same relationship is plotted, but stratified by the number of inspections made. The plot in panel (a) reflects the correlation between the sequence length and running maximum, as imposed by the mechanical rules under which the inspections are ordered. Conditioning on the inspection sequence removes this spurious correlation.

options. The order in which options are inspected is not even chosen by the consumers — it is mostly decided by the ranking algorithm. A correlation near zero therefore does not imply that the consumers do not optimize. A correlation near zero simply means that the ranking optimizes some other objective and that it is not sorting alternatives by the consumers' reservation values. This motivates our examination of the ranking process in Section 10.

5.4 Taking stock

Taken together, the first two tests are the ones that discriminate: the recall pattern and the conditional stopping evidence are both inconsistent with the canonical sequential-search rule and closer to a protocol in which the number of inspections is fixed in advance. The third is noninformative about consumer behavior but informative about platforms. Therefore, we will maintain fixed sample size of search as the chosen protocol, consistent with De los Santos et al. (2012) in a completely different market. This is not an innocuous choice, but is important to our ability to view our number of inspections as chosen prior to seeing their realization, as in K .

6 Model

6.1 Environment

A consumer i faces a purchase occasion, a finite set of restaurants indexed by j , and preferences where match value is additively separable into a vertical component common across consumers and an idiosyncratic component.

$$u_{ij} = \delta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma_{ij}^2) \text{ independent across } j. \quad (1)$$

The term δ_j captures what everyone agrees on, that bit of quality that you can at least partly observe from restaurant ratings and prices. The term ε_{ij} captures the idiosyncratic fit, that is, whether you as an individual consumer like this restaurant. That way, we can index each σ_{ij}^2 by both restaurant and customer.

This is precisely the starting point of both Stigler (1961) and Weitzman (1979). In their search models, all unsampled options share a common prior. The consumer is completely ignorant about each. Here, she is not. σ_{ij}^2 is weakly decreasing in the number of times that person i has previously ordered from seller j , so precision is an asset accumulated through consumption.

Another possibility is that the incumbent has already been consumed. In this case the match value u_{i0} is known to the consumer, and $\sigma_{i0}^2 = 0$ (and we write $v = u_{i0}$ for the value of the consumer's outside option), with the understanding that the v in our empirical setting includes options that are reachable via channels the browsing log does not record.

6.2 Search technology and timing

The consumer observes δ_j and her own priors, then selects an inspection set S with $|S| = K$, paying c per inspection. Inspecting j yields

$$s_{ij} = u_{ij} + \eta_{ij}, \quad \eta_{ij} \sim N(0, \sigma_\eta^2), \quad (2)$$

independent of everything else. Inspection can be informative only to a point: Photographs, menus and aggregate ratings tell the consumer something about a restaurant without revealing her own match with it.

Under normal updating, the posterior mean and variance are

$$\hat{\delta}_{ij} = \delta_j + \beta_{ij}(s_{ij} - \delta_j), \quad \beta_{ij} = \frac{\sigma_{ij}^2}{\sigma_{ij}^2 + \sigma_\eta^2}, \quad \tilde{\sigma}_{ij}^2 = \frac{\sigma_{ij}^2 \sigma_\eta^2}{\sigma_{ij}^2 + \sigma_\eta^2}. \quad (3)$$

Note that $\tilde{\sigma}_{ij}^2$ is strictly increasing in σ_{ij}^2 . You lower uncertainty somewhat, but not equally, with each examination. And remember, whatever product your buyer knew the least about before the sale will remain most uncertain after.

The consumer chooses the alternative (including the incumbent) with the highest pos-

terior mean. Consumption reveals the true u_{ij} and, in a subsequent period, the consumer will reorder from j if and only if $u_{ij} > \tau$, where τ is determined by her alternatives at that time.

Our research does not employ sequential search; we adopt a fixed sample size design. That is not a mere assumption of convenience; the data reject sequential design on three margins (see Section 5).

6.3 The search problem

Consider the symmetrical case when all alternatives have the same δ and σ . Given (δ, σ, v, K) , write $m \equiv \delta - v$ for the ex-ante advantage of a typical alternative over the incumbent. $Z_{(K)}$ denotes the maximum of K independent standard normal draws. The expected utility from inspecting K alternatives is

$$V(K) = v + \mathbb{E}[(m + \sigma Z_{(K)})^+] - cK, \quad (4)$$

using $\max\{v, \delta + \sigma Z\} = v + (m + \sigma Z)^+$. Define the gross marginal value of the K th inspection as $\Delta(K) \equiv V(K) - V(K-1) + c$.

Lemma 1. *For every $K \geq 1$,*

$$\Delta(K) = \sigma \int_a^\infty \Phi(z)^{K-1} [1 - \Phi(z)] dz, \quad a \equiv -m/\sigma.$$

Lemma 2. *$\Delta(K)$ is strictly decreasing in K . Hence V is strictly concave in K and the optimum satisfies $K^* = \max\{K : \Delta(K) \geq c\}$.*

Proofs of all results are in Appendix A.

6.4 Substitution between knowledge and search

Proposition 1. *Suppose $m \leq 0$, so that the incumbent is at least as good as the ex-ante mean alternative. Then $\Delta(K)$ is strictly increasing in σ for every K , the objective V is supermodular in (K, σ) , and the optimal search intensity $K^*(\sigma)$ is non-decreasing. Moreover $K^*(c)$ is non-increasing.*

This condition is also equivalent to the inequality $v \geq \delta$, where v indicates the restaurant a consumer returns to, while δ indicates an unfamiliar restaurant. Thus, $m \leq 0$ says that the restaurant a consumer returns to is at least as good as any unfamiliar restaurant. Empirically, the interesting case is that the consumer goes on exploring, rather than returning to the same place she already patronizes. This is essentially what happens in the comparative static above.

The immediate result of substitution is straightforward. For consumers whose current product set is completely dominated by alternatives with established names such that $\sigma_{i0} = 0$, they do not, as Proposition 1 predicts, search extensively because the marginal

private information gained from shopping around is small. Since uncertainty is reduced through either private information or shopping, consumers make use of the cheaper of the two technologies at the margin.

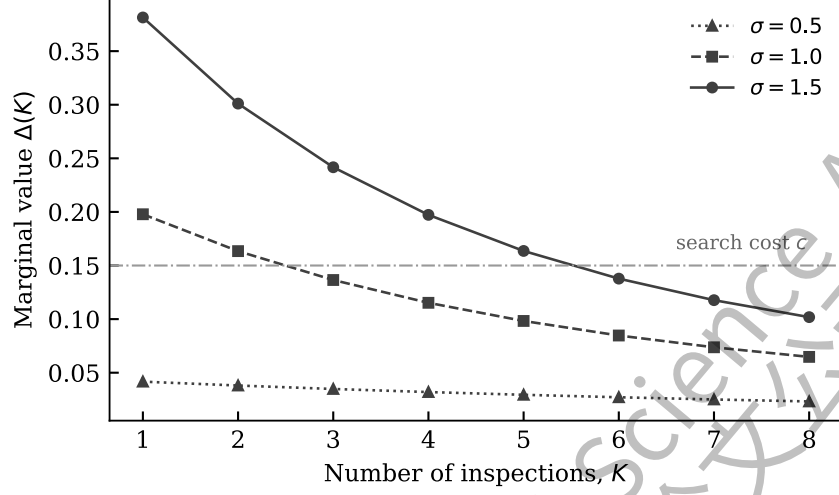


Figure 4: Marginal value of the K th inspection, computed by quadrature with $m = -0.5$. Reading across illustrates Lemma 2; reading between curves illustrates Proposition 1. The horizontal line marks an illustrative search cost; optimal K^* is where each curve crosses it.

6.5 Why searched matches are less durable

The next result contains the main empirical finding of this paper. It is important, however, to be precise about what the result establishes and does not.

We don't think this is an argument about winner's curse. Under proper Bayesian updating, $\mathbb{E}[u_{ij} \mid \hat{\delta}_{ij}] = \hat{\delta}_{ij}$, so that no "bias" exists in the expected value of the match resulting from simply selecting the alternative with the highest posterior mean. Our argument that the consumer chooses worse bargains than appeared at the onset must rely on the presence of systematic residual dispersion.

Proposition 2. *Let j^* be the chosen alternative, with posterior mean $\hat{\delta}^*$ and posterior variance $\tilde{\sigma}^2$, and suppose $\hat{\delta}^* > \tau$. Then*

$$\Pr(\text{repeat}) = \Phi\left(\frac{\hat{\delta}^* - \tau}{\tilde{\sigma}}\right),$$

which is strictly decreasing in $\tilde{\sigma}$ and therefore strictly decreasing in the prior variance σ^2 . Holding the posterior mean fixed, matches formed under diffuse priors are less durable than matches formed under precise priors.

Corollary 1. *Because $\tilde{\sigma}_{ij}^2$ is indexed by the consumer as well as the restaurant, conditioning on restaurant identity does not absorb the mechanism. The result should be robust after controlling for restaurants.*

The test in Corollary 1 provides our model’s most distinctive contribution. It allows us to distinguish two contrasting stories: In ours, search is beneficial for consumers; under the alternative story of pure adverse selection, it would be harmful, with restaurant fixed effects washing it out of the data. Our evidence (shown in Section 7) clearly rules out the latter explanation.

Corollary 2. *Two forces operate in opposite directions. Holding σ fixed, more inspections raise the posterior mean of the selected alternative, since $\mathbb{E}[\hat{\delta}^*] = \delta + \omega \mathbb{E}[Z_{(K)}]$ with ω the standard deviation of the posterior mean, and $\mathbb{E}[Z_{(K)}]$ is strictly increasing in K . Holding K fixed, higher σ lowers durability by Proposition 2. Because Proposition 1 implies that K and σ are positively associated in equilibrium, the sign of the observed relationship between search intensity and durability is not determined by theory.*

Corollary 2 is the formal reason this paper cannot answer the question “does searching help?” with a regression. The raw correlation between search and match durability confounds the return to inspecting more options with selection on how much the consumer already knew. Signing the first requires separating σ^2 from c , which requires the model.

6.6 Ranking design

Now let the platform choose which alternatives to display, and let the consumer inspect those displayed. The marginal contribution of displaying alternative j to a consumer whose best current option is worth v is the option value it creates:

$$MV(j) = \mathbb{E}[(\delta_j + \sigma_j z - v)^+], \quad z \sim N(0, 1). \quad (5)$$

Proposition 3. *Define $a_j = (v - \delta_j)/\sigma_j$. Then*

$$MV(j) = \sigma_j \varphi(a_j) + (\delta_j - v)[1 - \Phi(a_j)], \quad \frac{\partial MV}{\partial \delta_j} = 1 - \Phi(a_j), \quad \frac{\partial MV}{\partial \sigma_j} = \varphi(a_j) > 0.$$

The probability that j is chosen is exactly $1 - \Phi(a_j)$. A platform ranking by predicted purchase probability therefore ranks by $\partial MV/\partial \delta_j$ alone and omits the term $\sigma_j \varphi(a_j)$. A conversion ranking will be identical to the welfare-optimal ranking only if the constant σ_j is the same across all of your alternatives.

So a bit of accounting yields a neat result. The expected benefit from demonstrating the existence of restaurant j to a particular consumer can be decomposed into two pieces: The probability that she will patronize restaurant j , multiplied by its perceived advantage over her current alternative, plus a pure “option-value” term proportional to her uncertainty about j . A quantity that’s often calculated is expected purchase probability; it’s the first derivative from the first term, and it’s sufficient for the first term but does not contain any information about the second.

Corollary 3. *The neglected term $\sigma_j \varphi(a_j)$ is maximised, for given σ_j , at $a_j = 0$, that is at $\delta_j = v$. The alternatives a conversion-maximising ranking most under-displays are*

those that are simultaneously highly uncertain to the consumer and comparable in expected quality to her incumbent.

The most portable of these results is the set of alternatives under-shown by a conversion-optimised ranking. These are exactly those from which novel discovery will be made—the unfamiliar, about as good as what the consumer is currently buying. Alternatives which are much better or much worse than the incumbent are either promoted or suppressed, as should be the case under either objective. The wedge is where the other restaurants are: somewhere in the middle.

7 Search, Private Information, and Match Durability

7.1 Construction

The rest of the paper focuses on our main finding. A “discovery” is an order placed at a restaurant the consumer has not previously ordered from within the sample. It is *durable* if the consumer places another order at that same restaurant.

Assessing durability is tricky, because it requires the possibility of a repeat purchase. For this reason, we limit our attention to discoveries which are followed by at least three more orders by the same consumer. Of the 372,545 discoveries we analyze, 36.2% are durable.

7.2 The main result

Discoveries made without any recorded search are much more likely to become repeats than those that were the result of searches: 43.1% versus 30.1%. If we limit the comparison to a population of 45,821 people who make both kinds of discovery, our estimates show 43.8% versus 30.4% – a 13 percentage point difference.

This association is monotonic with durability: Durability is 43.1% when the restaurant receives no inspections and falls to 38.1%, 32.8%, 30.9%, 29.6% and 28.8% when the restaurant receives one through five inspections, respectively. It is 25.4% at six or more.

7.3 Interpretation, and what the result is not

A naive reading of Table 4 is that searching provides worse matches. It can’t be a reliable account – the model explains why.

In the model’s interpretation, browsing is most valuable where prior match uncertainty is high; Proposition 1 formalises the corresponding comparative static at the search-intensity margin. By equation (3), those alternatives retain the most residual uncertainty even after inspection. By Proposition 2, options chosen with high residual uncertainty are less likely to clear the threshold for a repeat purchase, not because they are worse in expectation but because they are more dispersed. The estimated relationship reflects selection on prior precision rather than a defect in search.

Table 4: Search intensity and the durability of discovered matches

	(1) Pooled OLS	(2) Consumer FE	(3) Restaurant FE	(4) Within-consumer difference
log(1 + inspections)	−0.0667 (0.0008) [−79.9]	−0.0651 (0.0013) [−49.0]	−0.0635 (0.0008) [−76.2]	
Searched – did not search				−0.1331 [−54.2]
Order-sequence control	Yes	Yes	Yes	—
Consumer fixed effects	No	Yes	No	Absorbed
Restaurant fixed effects	No	No	Yes	No
Observations	372,545	372,545	372,545	45,821

Notes. The dependent variable is an indicator that the discovered restaurant will be ordered from again later in the sample. Standard errors in parentheses, clustered at the consumer level. In brackets, we present the t -statistic for each. We have 105,235 clusters. Columns (1)–(3) include, as a control, an indicator for the consumer’s order-sequence number. The results in column (4) are based on comparing the durability of different types of discoveries – those made with and without a recorded inspection – within the population of consumers who make both kinds of discoveries. The unit of observation here is the consumer.

Two additional pieces of evidence favor this interpretation over the view that search makes consumers choose, *ceteris paribus*, observably worse restaurants.

Restaurant fixed effects barely matter. In column (3) of Table 4, restaurant fixed effects are included, which means that our estimate is driven entirely by variation across consumers who arrive at the same restaurant. The coefficient moves from -0.0667 to -0.0635 , a weakening of about 5%. Corollary 1 provides testable implications. Under the match-specific theory, σ_{ij}^2 varies from one pair of restaurant and customer to the next, and so cannot be absorbed by restaurant fixed effects. If consumers really did vary in restaurant quality preferences, then β should not be zero.

Observable quality is flat. The quality score is the average rating for the restaurant. In fact, quality is almost completely flat: at zero and 6+ inspections, mean quality is 4.696 and 4.692, with standard deviations stable over inspections at around 0.184 (the s.d. fluctuates between 0.180 and 0.185). Search does not sort toward observable quality: combined with Test 3 of Section 5, which finds essentially no quality ordering at all within the inspection sequence, we conclude that quality sorting is inactive along the vertical channel in the food delivery setting. All of the sorting power comes entirely from the elimination of idiosyncratic match uncertainty.

Both curves are drawn on the same chart in Figure 5, which provides perhaps the most visual demonstration of the paper’s findings – that the quality of what the consumer finds does not have to do with how intensely she examines the product, but the durability of what she finds does.

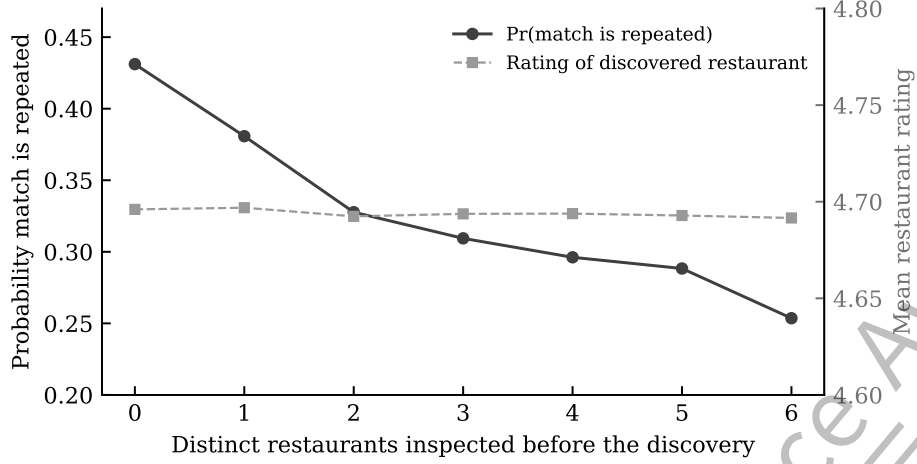


Figure 5: As search becomes more intensive, discovered matches become less durable, while the observable quality of the restaurant discovered stays essentially flat.

8 Heterogeneity: For Whom Is Search Worth Doing?

Proposition 1 implies that a consumer facing less prior uncertainty chooses fewer inspections. There is a second implication of search and private information being substitutes that is easier to test for directly: If they are substitutes, then the value of an additional unit of search to consumers should decline as a function of the level of private information available.

The facts in Table 2 are also exactly what we might expect. We find evidence for this substitution: the extra propensity to try a new restaurant that browsing buys falls monotonically as spending rises, from 10.6 percentage points in the lowest band to 6.0 percentage points in the highest band. The same substitution is visible in the estimated marginal returns: increasing the number of inspections from 0 to 6 raises the fraction of consumers searching by 16.1 percentage points in the lowest spending band and by only 9.6 percentage points in the highest.

We emphasize a null at these levels. The mean search is also nearly flat. Also, the decline in premiums reflects a genuine fall in the marginal benefit from search for consumers who spend the same amount of time and effort on product search.

8.1 Occasion-specific inertia

The model also implies that the cost of search should be reflected in meal patterns. The amount of time available for meal choice, on average, differs for each meal occasion, and the meal repetition rates do in fact vary systematically with available decision-making time. Meal repetition was 33.8% for breakfast, 31.2% for lunch, 27.7% in the afternoon, 27.8% at dinner and 27.2% late at night.

In Table 5 the rate gradients are close to monotone across meal occasions and across spending bands. The repeat rate ranges from 37.0% (for breakfast by the lowest spenders) to 23.7% (for dinner by the highest spenders).

Table 5: Repeat-order rates by meal occasion and consumer spending band

Meal occasion	Consumer spending band (RMB)				
	<29	29–36	36–49	49–65	≥65
Breakfast	0.370	0.344	0.341	0.336	0.317
Lunch	0.361	0.327	0.306	0.282	0.263
Afternoon	0.326	0.295	0.275	0.259	0.237
Dinner	0.333	0.296	0.276	0.254	0.237
Late night	0.310	0.283	0.260	0.254	0.238

Notes. Cell entries are the share of orders placed at a restaurant the consumer had ordered from previously. $N = 1,068,495$.

That the same person seems more habitual in the morning than in the evening is difficult to reconcile with a stable parameter, a habit preference, as is the standard approach of the state-dependence literature. It follows straightforwardly from a model in which inertia is an outcome of an ongoing search decision with occasion-specific costs.

8.2 Search on the left-hand side

This relationship places search on the right-hand side, where it is endogenous. A complement is the estimation of the determinants of search, using predetermined characteristics only.

$$\log(1+K_{it}) = \alpha_i + \beta_1 \text{Concentration}_{it} + \beta_2 \text{Breakfast}_t + \beta_3 (\text{Concentration}_{it} \times \text{Breakfast}_t) + \beta_4 k_{it} + \epsilon_i \quad (6)$$

where $\text{Concentration}_{it}$ is the Herfindahl index of consumer i 's own order history strictly before t , a measure of how concentrated her past patronage has been.

Table 6 raises some methodological points we feel are important enough to mention in the main text. Column (1) uses the obvious and natural measure of familiarity: the probability that a consumer had previously ordered at a restaurant. It yields a large negative coefficient. As it happens, however, that measure is not predetermined; consumers' and restaurants' subsequent choices of how much searching to do are correlated. In column (2), we replace it with the degree of concentration of restaurants visited by the consumer in the past. The sign of the coefficient changes. Search is positively correlated with how concentrated the consumer's restaurant patronage is.

An interaction with breakfast, which an obviously simple attention-saving story would predict to be negative, is positive and significant in column (1) but statistically zero in column (2). We discuss this as it does not support the original interpretation.

This effect holds across all specifications. Breakfast search is around 0.08 log points less than on other occasions within the same consumer. The t -statistics lie between 15 and 21. This is the variation that would be used to define the cost distribution in a structural version.

Table 6: Determinants of search intensity

	(1) Familiarity with the restaurant eventually chosen	(2) Concentration of prior order history
Familiarity measure	−0.0509 (0.0011) [−44.7]	+0.0679 (0.0024) [+28.1]
Breakfast	−0.0867 (0.0042) [−20.5]	−0.0820 (0.0054) [−15.2]
Familiarity × Breakfast	+0.0125 (0.0023) [+5.5]	−0.0088 (0.0090) [−1.0]
Consumer fixed effects	Yes	Yes
Observations	1,067,425	1,067,425

Notes. Dependent variable: $\log(1 + \text{distinct inspections})$. Standard errors are presented in parentheses, clustered at the consumer level; the statistics in brackets are t -statistics, computed with 200,000 clusters. Both columns include a control for the consumer's order-sequence index. In column (1), familiarity is measured for the restaurant the consumer eventually chooses; that measure is not predetermined, since the restaurant identity is determined jointly with search. In column (2), concentration is calculated just before the purchase to avoid endogeneity bias. If we had used the concentration of the consumer's complete purchase history, the sign of the coefficient on concentration in column (2) would have reversed.

9 Robustness

9.1 Inference

The standard errors on all estimates below in this paper are based on cluster estimation at the consumer level. This treatment is necessary because the orders of a given consumer are manifestly correlated with each other — a given consumer has stable tastes, a stable address, and a stable meal schedule — and are not independent observations as if the orders came from an infinite population of buyers. There are 105,235 consumer clusters in the durability sample, 200,000 consumer clusters in the full order sample, and the asymptotic approximation to the variance is certainly not in doubt.

Clustering leads to wider intervals, as can be seen for search intensity in Table 4: in the consumer fixed-effects specification the standard error rises from 0.0011 to 0.0013, with a corresponding decrease in the t -statistic from 58.3 to 49.0. On a relative basis, the biggest jump in magnitude was for the breakfast indicator, whose t -statistic fell from 23.8 to 20.5. However, no conclusion of the paper depends on the distinction.

9.2 Selection into the browsing log

If our results are driven by the selection process into the browsing log, the analysis is more troubling. For example, the results would be misleading if the session types in which search and purchase events were both logged differed from those in which they were not.

We assess this directly by looking at the fraction of a consumer’s orders that have some inspection associated with them, at the consumer level. We split consumers into deciles based on how complete the log is for their activity, and re-estimate the durability specification within each decile.

Table 7: Durability effect by decile of logging completeness

Decile	1	2	3	4	5	6	7	8	9
Mean logging share	0.09	0.33	0.46	0.55	0.60	0.68	0.77	0.85	0.96
$\log(1 + \text{inspections})$	-0.065 (0.004)	-0.072 (0.004)	-0.076 (0.002)	-0.069 (0.005)	-0.069 (0.003)	-0.065 (0.003)	-0.065 (0.002)	-0.057 (0.003)	-0.042 (0.003)
Observations	81,262	31,566	59,637	14,442	37,871	36,994	43,758	30,048	36,967

Notes. Consumers are split into deciles by the fraction of their orders that have at least one logged inspection. Within each decile we estimate a linear probability model of match durability on $\log(1 + \text{inspections})$ and the order-sequence index. Standard errors clustered at the consumer level are in parentheses. We omit deciles with fewer than 2,000 discoveries. The full-sample estimate is -0.0667 .

In Table 7 we continue to find a strong negative and significant effect of search intensity in every decile of logging completeness (-0.076 to -0.042). Thus, the effect is not an artifact of logging completeness or a subgroup of the data.

Still, there appears to be some attenuation. In the highest logging-completeness decile, for which over 96% of consumers’ orders are logged, the coefficient is -0.042 , about a

third below its full-sample value. The point estimate may be read in one of two ways: Either consumers in the top decile are heavy browsers for whom search has essentially no marginal value—an interpretation the model would anticipate—or else the estimated elasticity reflects unlogged search in other observations being incorrectly recorded as zero searches. The latter possibility, which we cannot rule out with the available data, requires us to view our full-sample coefficient as an upper limit and -0.042 as a lower limit on its magnitude. The qualitative conclusion that “searched” discoveries are markedly less durable than unsearched ones seems robust to this re-interpretation.

9.3 Alternative definitions of durability

A discovery can only be classified as durable or not if the consumer is observed again after it, so the baseline requires at least three subsequent orders by the same consumer. Table 8 varies that requirement from two to five.

Table 8: Sensitivity to the observation-window requirement

Minimum subsequent orders required	2	3	4	5
$\log(1 + \text{inspections})$	-0.0663 (0.0007)	-0.0667 (0.0008)	-0.0669 (0.0009)	-0.0674 (0.0011)
Durability, no search	0.402	0.431	0.455	0.476
Durability, searched	0.273	0.301	0.324	0.344
Difference	-0.130	-0.131	-0.131	-0.133
Observations	475,129	372,545	297,077	239,442

Notes. Column 3 is the specification reported in Table 4. Consumer-clustered standard errors in parentheses.

Point estimates of the coefficient vary from -0.0663 to -0.0674 as the threshold moves from two to five, while the raw difference in discoveries searched for versus not searched for lies between 13.0 and 13.3 percentage points. Durability is mechanically increasing in the window; that is, with a longer duration the chance of observing a repeat increases because the window has afforded more opportunity to recur, but the effect we care about, the relative difference, is constant.

9.4 Functional form

All of our models use linear probability specifications. The dependent variable has a mean of 0.362. Without a great deal of mass at the boundaries, a linear approximation is probably not misleading, and it allows for a sufficiently high dimension of fixed effects, which Corollary 1 requires.

We nonetheless re-estimate the baseline model on the logit scale. The coefficient on $\log$ inspections, -0.3064 (0.0038), implies an average partial effect of -0.0685 . That is -0.0685 against -0.0667 from the linear model—the difference is under 3% of the estimate. Nothing in the paper turns on the choice.

10 Implications for Ranking Design

10.1 The wedge

Proposition 3 establishes that the value of displaying an alternative to the incumbent has two parts, only one of which is captured by the commonly predicted increase in purchase probability. Corollary 3 shows that the distortion is largest when the alternative is of uncertain quality to the consumer but relatively similar to her current option.

Test 3 in Section 5 provides indirect evidence that this wedge is operative. If inspection order were generated by a reservation-value index, then inspection position would be correlated with a function of both quality and uncertainty. However, it is essentially uncorrelated ($r = -0.00851$) with quality. So the ranking faced by consumers in the market is not the ranking that ultimately determines welfare.

10.2 What would be required to quantify it

Signing the wedge requires only the model. To measure the wedge requires estimation of the distribution of search costs, the functional mapping of observable familiarity to prior heterogeneity in value, and the precision of the information in each search: The most straightforward approach follows the tradition of the literature on fixed-sample search models, calculating a likelihood contribution for a given session as the probability of the observed number of inspections multiplied by the probability that the observed purchase is the utility maximizer among the inspected options.

$$L_{it} = \Pr(K = K_{it}^{\text{obs}} | c, \sigma, \delta) \times \Pr\left(j_{it}^{\text{obs}} = \arg \max_{j \in S_{it}} u_{ij}\right).$$

The pattern of identifications comes from three sources. The extent of variation of inspections per consumer over different meal occasions allows us to estimate the search cost distribution. The consumer effect observed in Table 6 is both large and precisely estimated. Also, the durability between matches that were formed with and without search gives us insight into the prior variance, via Proposition 2. Finally, the distribution of inspection counts and of matching outcomes provides information on the remaining scale parameters.

Then, three display rules are compared: the display rule we observe, the display rule that was associated with the most probable future purchase, and the display rule corresponding to the highest reservation-value index. The interesting results relate to the difference between the second and third of these rules, with respect to both expected match value and search time. Since the cost of one minute of consumer attention is not identified, the break-even bands across the range of possible time cost functions are probably more informative than a point estimate.

10.3 A second dimension worth exploring

The two categories (food and delivery quality) are reported separately in the restaurant catalogue. These are only weakly correlated (correlation = 0.306 among rated restaurants). One consequence of the ranking is that it brings out one dimension at the expense of the other; so this gives us a deeper question to ask, rather than the previous question of how much exploration we should be inducing on a platform: Which attribute do you want it to show?

11 Estimating the Model

The previous section set out what would be needed to put a number on the wedge. We now do it.

11.1 What makes the model estimable

The obstacle is that search costs are never observed. What rescues us is the stopping rule itself. By Lemma 2 the consumer inspects until the marginal value of another look falls below its cost, so $K^* = \max\{K : \Delta(K) \geq c\}$ with Δ declining in K . Seeing how many restaurants somebody opened therefore does not tell us her search cost, but it does tell us which interval her cost falls in:

$$K = 0 \iff c \geq \Delta(1), \quad (7)$$

$$K = k \iff \Delta(k+1) \leq c < \Delta(k), \quad k \geq 1. \quad (8)$$

Every browsing session is thus an interval-censored reading on that consumer's cost of looking. Assuming c is lognormal, with its location allowed to shift across meal occasions, the likelihood follows in closed form once Δ is evaluated by the quadrature in Lemma 1. No data beyond the inspection counts is required.

One normalisation is unavoidable. Lemma 1 gives $\Delta(K) = \sigma \cdot g(K, a)$ with $a = -m/\sigma$, so a doubling of σ and of c together leaves the inspection distribution unchanged. We therefore fix $\sigma = 1$ and report every cost in units of one standard deviation of match value. The ratio a , which measures how strong the incumbent is relative to the spread of match values, is not affected by this and is estimated from the *shape* of the distribution.

11.2 Estimates

Three things in Table 9 are worth pausing on.

The median cost of opening one more restaurant page is 0.44 standard deviations of match value. That is a lot. Putting it another way, a consumer would rather accept a meal that suits her almost half a standard deviation worse than click through to one more listing. It also explains something the raw data makes obvious but which is easy to wave away: just over half of all orders involve no recorded browsing at all.

Table 9: Structural estimates of the search-cost distribution

Parameter	Estimate	Std. error
a strength of the incumbent	+0.0068	0.0042
μ log median search cost	-0.8172	0.0076
$\log s$ dispersion of log costs	+0.3418	0.0030
<i>Shift in μ by meal occasion (breakfast omitted)</i>		
Lunch	+0.0072	0.0058
Afternoon	-0.1314	0.0071
Dinner	-0.1064	0.0060
Late night	-0.1864	0.0080
Implied median cost, breakfast	0.442	
Implied median cost, late night	0.367	
Log-likelihood	-1,904,019	
Observations	1,068,495	

Notes. Maximum likelihood on the distribution of distinct inspections per order, top-coded at twelve. Costs are expressed in units of one standard deviation of match value under the normalisation $\sigma = 1$. Standard errors come from the numerical Hessian.

Search costs are highest at breakfast and lowest late at night, with the late-night median running about 17 percent below the breakfast one. Nothing in the estimation imposed that ordering. It reproduces the occasion pattern found earlier in Table 6, which is reassuring, because those two results come from different moments of the data.

Finally, the dispersion parameter is large, $s = 1.41$. Search costs differ enormously from one consumer and occasion to the next. That heterogeneity is what lets a single distribution generate both an enormous mass at zero inspections and a tail that runs out past twelve.

11.3 Does the model fit?

The estimation uses seven parameters to match thirteen moments, so fit is a genuine test rather than an accounting identity.

Figure 6 shows the answer is yes, and not only where it is easy. The model puts 51.5 percent of orders at zero inspections against 51.6 observed, and it tracks the decline through the middle of the distribution and the mass beyond twelve. The one visible miss is at one versus two inspections, where the model predicts too many single inspections and too few pairs. A smooth cost distribution cannot easily produce a consumer who opens a second listing out of habit, and we suspect that is what the gap represents.

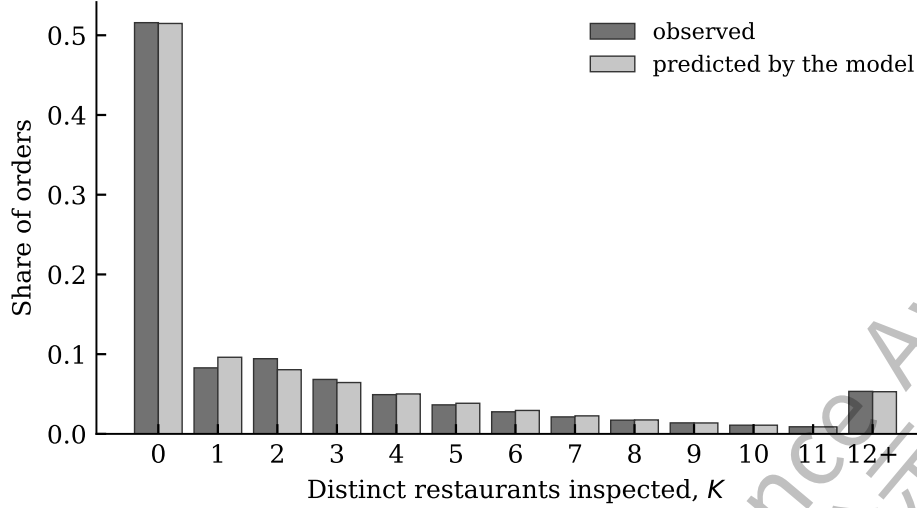


Figure 6: Observed and predicted distribution of inspection counts. Predicted shares aggregate across meal occasions using the observed occasion weights.

12 The Wedge, Quantified

12.1 Setting up the comparison

With the cost distribution in hand, the display rule becomes something we can vary. We simulate consumers facing forty candidate restaurants whose vertical quality is drawn standard normal. A tenth of them are familiar and carry little prior uncertainty; the rest are unfamiliar and carry a great deal. Each consumer draws her own search cost from the estimated distribution and chooses how many restaurants to inspect using the model's own stopping rule.

We then hold the number of inspections fixed and change only *which* restaurants occupy those slots. This matters for interpretation: the exercise is not about making consumers search more, which would cost them time. It is about what they are shown for the same amount of looking. Three rules are compared: a random ordering, as a floor; a ranking by purchase probability $1 - \Phi(a_j)$, which is what a recommender trained on conversion implements; and a ranking by the welfare index $MV(j)$ of Proposition 3.

12.2 Results

Reordering the same set of slots by the welfare index rather than by purchase probability raises expected match value by 6.9 percent and consumer surplus by 8.5 percent. The consumer spends no additional time searching; she is simply shown a different set of restaurants.

The gap between random ordering and conversion ranking, at 45.7 percent, is worth reporting alongside it. Most of what a ranking can accomplish, a conversion ranking already accomplishes. The distortion we identify is the residual, and it would be wrong to read Table 10 as saying that commercial rankings are badly broken. They are not.

Table 10: Welfare under alternative display rules

Display rule	Match value	Search cost	Surplus	Mean K	vs. conversion
Random order	1.358	0.470	0.888	4.59	-45.7%
Conversion ranking	2.499	0.470	2.029	4.59	—
Reservation ranking	2.671	0.470	2.201	4.59	+6.9%

Notes. All values in units of match-value standard deviations. The final column gives the change in expected match value relative to conversion ranking; in consumer surplus the corresponding gain is 8.5 percent. Search intensity is held fixed across rules, so the comparison isolates composition. Baseline calibration: a tenth of alternatives familiar, prior standard deviation of 1.0 on the rest.

They are slightly and systematically tilted, in a direction that happens to matter for discovery.

12.3 How much does the calibration drive this?

Two quantities are not identified by inspection counts: how much of a consumer’s choice set she already knows, and how uncertain she is about the rest. Rather than pick values and present a single number, we report the counterfactual across a grid.

Table 11: Gain from reservation ranking across calibrations

Prior SD on unfamiliar alternatives	Share of alternatives already familiar		
	0.05	0.10	0.20
0.75	2.1%	3.8%	6.6%
1.00	3.8%	6.6%	11.2%
1.25	5.1%	9.3%	16.0%
1.50	6.1%	11.4%	20.9%

Notes. Percentage gain in expected match value from reservation ranking relative to conversion ranking. The baseline reported in Table 10 is shown in bold.

The gain runs from 2.1 to 20.9 percent across the grid. It rises in both directions, and the reason is visible in Proposition 3: the neglected term is $\sigma_j \varphi(a_j)$, which grows when consumers are more uncertain and when more of their choice set is familiar, since familiarity raises the incumbent value against which everything else is judged. So the direction of the result is robust and its size is not. We report the range rather than pretending to a precision the data cannot support.

12.4 Where the distortion falls

Corollary 3 already located it: the neglected term peaks where $\delta_j = v$. The restaurants a conversion ranking most under-shows are the ones a consumer knows least about that are also roughly as good as the place she already orders from.

This connects back to something noted in Section 4 and left hanging there. Concentration in these districts is negligible by any standard measure, and yet the median

consumer uses three restaurants in three weeks. What limits effective competition here is not the number of sellers but the scarcity of attention, and the ranking is what allocates attention. A platform maximising conversion allocates it in a way that is close to what consumers would choose for themselves, but not the same, and the difference falls precisely on the restaurants they would otherwise have discovered.

13 Limitations

Here we sum the claims the paper cannot support.

The browsing log is admittedly incomplete. The restaurant actually ordered from appears in the recorded browsing sequence only 19.4% of the time across all multi-inspection sessions. Results reported in the rest of the paper thus describe search behavior as it appears on the platform’s recommendation surfaces. Nevertheless the durability effect is observed in every decile of completeness in Table 7, and cannot be explained by behavior concentrated at an extreme tail of the distribution. Across completeness deciles the durability effect is bounded between -0.076 and -0.042 , and although we have not modeled the unobserved channel, we suspect these estimates would be largely unchanged had we instead recorded search queries, created favorites, or even offline restaurant knowledge; to truly do so, one simply must include the unlogged channel in the estimated set of alternatives, an approach we have not taken.

Search is not imposed. Consumers who want to try something new presumably browse more. Nothing in our design rules out this simultaneity, and the resulting coefficients on search intensity are merely associations. Our results are thus, at most, indicative of interactions. The gradients in Section 8 are more robust than the levels, since there is no obvious reason why simultaneity bias should vary systematically across bands, but of course that’s an argument rather than an identification strategy.

But remember: the spending bands are not income. They are just categories the service provider uses to make sense of the spending patterns of different types of consumers (and we’re missing some of the data here).

Estimates are approximate. Prices in the data are reported as ranges rather than exact order values, and item prices do not always add up to reported order prices.

We estimate the model in Section 11, but two parameters governing prior uncertainty are calibrated rather than identified, so the size of the wedge in Section 12 is reported as a range rather than a point. Keep in mind, finally, that the ranking exercise in Section 10 is not a diagnosis.

It all happens in one city and over a span of three weeks, so the results cannot be generalised beyond Beijing in March 2021.

14 Conclusion

Platforms rank alternatives to maximise the probability that a consumer transacts. This paper has argued that the objective is systematically misaligned with consumer welfare and characterised the misalignment precisely: predicted purchase probability is the derivative of welfare with respect to expected quality, and carries no information about the option value created by showing a consumer something she is uncertain about. Because that option value is largest for alternatives similar in expected quality to what the consumer already buys, the restaurants a conversion-optimised ranking most under-displays are exactly those from which discovery would come.

The empirical work supports the model’s central premise — that prior precision differs across alternatives and that search is directed at what consumers know least — through a result that initially looks like a puzzle. Restaurants found through browsing are substantially less likely to become regular choices than restaurants reached without it, by 13.3 percentage points within consumer. The effect survives restaurant fixed effects almost intact and is unaccompanied by any variation in the observable quality of what consumers find. Within the model this is not a failure of search but a consequence of its being deployed where uncertainty is greatest.

That finding carries a methodological warning. Because search intensity and prior uncertainty are positively linked in equilibrium, the sign of their reduced-form association with match quality is not determined by theory. Reading the returns to search off a regression of outcomes on search intensity recovers a mixture of two effects with opposite signs. Estimating the model separates them, and doing so suggests that reordering a display to account for option value would raise expected match quality by around 7 percent at our baseline calibration, with a range of 2 to 21 percent across the parameters we cannot pin down.

Three extensions seem worth pursuing. The first is to model the unlogged search channel explicitly, which would convert our bounds into estimates. The second exploits a feature of this setting we have set aside: food quality and delivery quality are separately reported and only weakly correlated, so a platform choosing what to make salient faces a composition problem as well as a quantity problem. Which dimension a ranking should surface strikes us as at least as interesting as how much discovery it should induce. The third is to take the wedge to a setting with experimental variation in rankings, as in Ursu (2018), where the counterfactual could be validated rather than simulated.

References

- Bergemann, D. and Bonatti, A. (2024). Data, competition, and digital platforms. *American Economic Review*, 114(8):2553–2595.
- Brown, J. R. and Goolsbee, A. (2002). Does the internet make markets more competitive? evidence from the life insurance industry. *Journal of Political Economy*, 110(3):481–507.
- De los Santos, B., Hortaçsu, A., and Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, 102(6):2955–2980.
- Dinerstein, M., Einav, L., Levin, J., and Sundaresan, N. (2018). Consumer price search and platform design in internet commerce. *American Economic Review*, 108(7):1820–1859.
- Dubé, J.-P., Hitsch, G. J., and Rossi, P. E. (2010). State dependence and alternative explanations for consumer inertia. *RAND Journal of Economics*, 41(3):417–445.
- Jin, G. Z. and Leslie, P. (2003). The effect of information on product quality: Evidence from restaurant hygiene grade cards. *Quarterly Journal of Economics*, 118(2):409–451.
- Sorensen, A. T. (2000). Equilibrium price dispersion in retail markets for prescription drugs. *Journal of Political Economy*, 108(4):833–850.
- Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, 69(3):213–225.
- Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, 37(4):530–552.
- Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica*, 47(3):641–654.

致谢

一、研究背景与选题来源

本研究源于作者本人与家人朋友使用外卖平台的生活经验。长期使用过程中，作者注意到平台推荐页面经常优先展示已经熟悉或反复下单的商家，而真正可能带来新发现的商家并不总处于显眼位置。由此产生了本文的核心问题：平台以“成交概率”为排序目标时，是否与消费者希望发现更合适选择的目标完全一致？在查阅消费者搜索（consumer search）、平台设计（platform design）与重复购买等文献后，作者逐步将这一观察转化为可建模、可检验的经济学问题，并形成本文关于“成交”与“发现”之间权衡的研究框架。

二、作者与指导老师的分工

作者独立完成的研究工作。研究问题、模型核心设定、数据获取与清洗、变量构造、全部实证分析、结构式估计、反事实模拟及论文初稿均由作者独立完成。作者自行获取美团 Meituan Takeout Recommendation Dataset（Zenodo 公开数据集，CC BY 4.0），编写全部数据处理与计算代码，并独立完成搜索协议检验、固定效应回归、稳健性检验、结构参数估计与排序反事实分析。论文中的研究判断、结果解释及最终责任均由作者本人承担。定稿阶段曾使用生成式 AI 进行有限的校对与表达检查，具体情况见第四部分。

指导老师提供的帮助。指导老师主要在两方面提供学术指导：一是提示消费者搜索与平台排序领域的重要文献，帮助作者更准确地定位研究；二是就模型证明的严谨性与比较静态条件提出修改意见。指导老师未参与数据处理、程序编写、实证估计、反事实模拟或论文代写。本次指导为无偿指导，未收取任何形式的报酬。

三、研究过程中遇到的困难及解决经过

第一，最初的理论解释被数据否定。作者曾尝试用“预算约束下的风险管理”解释低支出消费者的重复购买，但预先检验发现数据中的方向与模型预测相反，因此放弃原模型并重新构建以搜索与先验不确定性为核心的框架。这一过程促使作者坚持先检验可证伪预测，再决定是否继续发展理论。

第二，熟悉度变量的构造暴露出内生性问题。最初使用“最终选择商家的历史下单次数”衡量熟悉度，但后来意识到最终选择与搜索强度是同时决定的。改用严格前定的历史订单集中度后，相关系数由约 -0.051 变为约 $+0.068$ 。因此正文同时报告两种设定，以说明变量时序与识别假设的重要性。

第三，浏览日志并不完整。多次浏览会话中，最终下单商家仅有 19.4% 出现在记录的浏览序列中。作者因此将研究对象明确限定为“平台推荐界面上可观测的搜索行为”，并按日志完整度分组重复估计，以检验核心结果是否由记录缺失驱动。

第四，搜索成本无法直接观测。作者利用模型的停止规则，将每次会话的浏览次数转化为搜索成本的区间删失信息，并据此构造极大似然估计，使结构式估计能够在现有数据条件下实施。

四、AI 工具使用情况

本文在研究主体完成后、提交前的定稿校对阶段使用了生成式 AI，使用范围限定如下。

Claude (Opus 5, High 模式)。主要进行了两类集中式审阅。第一类为数学与公式一致性检查：作者要求其通读论文并核对模型、公式与证明是否前后一致。该轮审阅指出了 Proposition 3 证明中的一个链式法则符号书写错误、搜索回归中的符号命名冲突，以及若干文字或数字表述问题；作者随后逐项核对原始推导和数据后决定是否修改。第二类为学术英语诊断：Claude 对全文英语表达进行审阅，列出难以解析的句子、第二人称使用、英美拼写混用、代词一致性以及口语化缩写等问题；该轮主要提供修改清单，具体改写由作者自行完成。Claude 未用于提出研究问题、构建核心理论、选择数据、编写实证代码、运行估计或生成研究结果。

总体而言，AI 的使用集中在论文完成后的校对与披露整理阶段，并非持续参与研究过程。所有实质性研究内容、代码、结果与结论均由作者独立完成和核验。

五、致谢

感谢指导老师在文献与证明方面给予的指点。感谢 Meituan（美团）与数据集发布者以开放许可提供数据，使本研究成为可能。文中一切错误与不足，均由作者本人负责。

A Proofs

Proof of Lemma 1

Let $h(z) = (m + \sigma z)^+$, which is non-decreasing, piecewise linear, satisfies $h(-\infty) = 0$, and has $h'(z) = \sigma \mathbf{1}\{z > a\}$ almost everywhere with $a = -m/\sigma$. For any random variable X with cdf F , integration by parts gives

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h'(z)[1 - F(z)] dz.$$

The maximum of K independent standard normals has cdf $\Phi(z)^K$, so

$$\mathbb{E}[h(Z_{(K)})] = \sigma \int_a^{\infty} [1 - \Phi(z)^K] dz.$$

Differencing at K and $K - 1$ and using $\Phi^{K-1} - \Phi^K = \Phi^{K-1}(1 - \Phi)$ yields the stated expression. ■

Proof of Lemma 2

On the domain of integration $\Phi(z) \in (0, 1)$, so $\Phi(z)^{K-1}$ is strictly decreasing in K pointwise and the integrand of Lemma 1 inherits this property. Hence $\Delta(K)$ is strictly decreasing. Since the marginal cost of an inspection is the constant c , the objective V is strictly concave in K and the optimum is characterised by the stated stopping rule. ■

Proof of Proposition 1

Write $a = -m/\sigma$, so $a \geq 0$ under the maintained condition $m \leq 0$, and note $\partial a/\partial \sigma = m/\sigma^2 = -a/\sigma$. Differentiating the expression in Lemma 1 with respect to σ , applying Leibniz's rule to the σ -dependent lower limit,

$$\frac{\partial \Delta(K)}{\partial \sigma} = \int_a^{\infty} \Phi^{K-1}(1 - \Phi) dz - \sigma \Phi(a)^{K-1}[1 - \Phi(a)] \cdot \frac{\partial a}{\partial \sigma} = \int_a^{\infty} \Phi^{K-1}(1 - \Phi) dz + a \Phi(a)^{K-1}[1 - \Phi(a)].$$

The first term is strictly positive because the integrand is strictly positive on the interior of the domain. The second is non-negative precisely because $a \geq 0$. Hence $\partial \Delta(K)/\partial \sigma > 0$.

Since $\Delta(K) = V(K) - V(K - 1) + c$ is increasing in σ for every K , the objective V has increasing differences in (K, σ) , and Topkis's monotone selection theorem gives that $K^*(\sigma)$ is non-decreasing. For the comparative static in c , note that $\partial^2 V/\partial K \partial c = -1 < 0$, so V has strictly decreasing differences in (K, c) and $K^*(c)$ is non-increasing. ■

Proof of Proposition 2

Conditional on the inspected signal, the true match value satisfies $u_{ij^*} \mid s_{ij^*} \sim N(\hat{\delta}^*, \tilde{\sigma}^2)$. A repeat occurs if and only if $u_{ij^*} > \tau$, so

$$\Pr(\text{repeat}) = 1 - \Phi\left(\frac{\tau - \hat{\delta}^*}{\tilde{\sigma}}\right) = \Phi\left(\frac{\hat{\delta}^* - \tau}{\tilde{\sigma}}\right).$$

Differentiating,

$$\frac{\partial \Pr(\text{repeat})}{\partial \tilde{\sigma}} = \varphi\left(\frac{\hat{\delta}^* - \tau}{\tilde{\sigma}}\right) \cdot \left(-\frac{\hat{\delta}^* - \tau}{\tilde{\sigma}^2}\right) < 0 \quad \text{whenever } \hat{\delta}^* > \tau.$$

Finally, from equation (3), $\partial \tilde{\sigma}^2 / \partial \sigma^2 = \sigma_\eta^4 / (\sigma^2 + \sigma_\eta^2)^2 > 0$, so $\tilde{\sigma}$ is strictly increasing in the prior variance. Composing the two derivatives gives the result. ■

Proof of Corollary 1

Immediate from the indexing in equation (1). The term δ_j is common across consumers and is absorbed by a restaurant fixed effect; the term σ_{ij}^2 varies across consumers at a given restaurant and is not. ■

Proof of Proposition 3

Substituting $a_j = (v - \delta_j) / \sigma_j$ into equation (5) and using the standard normal partial-expectation identity $\mathbb{E}[(z - a)^+] = \varphi(a) - a[1 - \Phi(a)]$,

$$MV(j) = \sigma_j \mathbb{E}[(z - a_j)^+] = \sigma_j \varphi(a_j) - \sigma_j a_j [1 - \Phi(a_j)] = \sigma_j \varphi(a_j) + (\delta_j - v) [1 - \Phi(a_j)],$$

since $\sigma_j a_j = v - \delta_j$.

For the derivative with respect to σ_j , note that $v - \delta_j$ does not depend on σ_j , that $\partial a_j / \partial \sigma_j = -a_j / \sigma_j$, and that $\varphi'(a) = -a\varphi(a)$. Then

$$\frac{\partial MV}{\partial \sigma_j} = \varphi(a_j) + \sigma_j \varphi'(a_j) \left(-\frac{a_j}{\sigma_j}\right) + (v - \delta_j) \varphi(a_j) \left(-\frac{a_j}{\sigma_j}\right) = \varphi(a_j) + a_j^2 \varphi(a_j) - a_j^2 \varphi(a_j) = \varphi(a_j).$$

For the derivative with respect to δ_j , with $\partial a_j / \partial \delta_j = -1 / \sigma_j$,

$$\frac{\partial MV}{\partial \delta_j} = \sigma_j \varphi'(a_j) \left(-\frac{1}{\sigma_j}\right) + [1 - \Phi(a_j)] + (v - \delta_j) \varphi(a_j) \left(-\frac{1}{\sigma_j}\right) = a_j \varphi(a_j) + [1 - \Phi(a_j)] - a_j \varphi(a_j) = 1 - \Phi(a_j)$$

where the cancellation uses $(\delta_j - v) / \sigma_j = -a_j$.

Finally $\Pr(j \text{ chosen}) = \Pr(\delta_j + \sigma_j z > v) = \Pr(z > a_j) = 1 - \Phi(a_j)$. Along an iso-MV curve in (δ, σ) space the slope is $d\sigma/d\delta = -[1 - \Phi(a)]/\varphi(a) < 0$, so alternatives that are welfare-equivalent to display are strictly ordered by conversion whenever they differ in σ .

Conversion ranking is therefore welfare-optimal only in the degenerate case of constant σ . ■

Proof of Corollary 3

The neglected term is $\sigma_j \varphi(a_j)$. For fixed σ_j this is maximised where φ is maximised, namely at $a_j = 0$, which by the definition of a_j occurs at $\delta_j = v$. ■

B Additional Tables

Table 12: Order composition across the day

	Breakfast	Lunch	Afternoon	Dinner	Late night
Orders	87,540	468,743	115,423	328,154	68,635
Median hour (Beijing)	8	11	15	18	21
Share of orders <29 RMB	0.535	0.468	0.370	0.349	0.359
Share of orders $\geq$ 65 RMB	0.065	0.163	0.208	0.248	0.190
Mean distinct inspections	2.53	2.45	3.02	2.86	3.04
Exploration rate	0.662	0.688	0.723	0.722	0.728

Notes. Meal occasion is the platform's own classification. Exploration is an indicator for ordering from a restaurant the consumer has not previously used.

Table 13: Market concentration by commercial district

District	Orders	Restaurants	Herfindahl index
0	241,422	2,497	18
1	65,489	1,278	53
2	61,370	1,235	46
3	104,537	1,465	33
4	55,027	1,094	47
5	66,781	1,420	38
6	174,142	1,983	24
7	32,941	1,013	54
8	86,785	1,346	44
9	82,944	1,952	27
10	97,057	1,560	39

Notes. Herfindahl indices are computed over restaurant shares of district orders and reported on a scale to 10,000. The ten largest restaurants account for between 6.2% and 15.8% of orders depending on the district.

Table 14: Marginal returns to inspection by consumer spending band

Distinct inspections	0	1	2	3	4	5	6
Spending band <29	0.600	0.607	0.677	0.709	0.723	0.740	0.761
Spending band 36–49	0.664	0.673	0.728	0.751	0.765	0.783	0.796
Spending band ≥ 65	0.718	0.716	0.758	0.782	0.787	0.794	0.814

Notes. Cell entries are exploration rates. Moving from zero to six inspections raises exploration by 16.1 percentage points in the lowest band and 9.6 in the highest.