You Can Lead a Horse to Water: Spatial Learning and Path Dependence in Consumer Search*

Charles Hodgson
Stanford University

Gregory Lewis
Microsoft Research and NBER

February 12, 2019

Abstract

We develop a model of search by imperfectly informed consumers with unit demand. The innovation is that consumers learn spatially: sampling the payoff to one product causes them to update their payoffs about all products that are nearby in some attribute space. Search is costly, and so consumers face a trade-off between "exploring" far apart regions of the attribute space and "exploiting" the areas they already know they like. Learning gives rise to path dependence, as each new search decision depends on past experiences through the updating process. We present evidence of these phenomena in data on online camera purchases, showing that the search paths and eventual purchase decisions depend substantially on whether the past items searched were surprisingly good or bad. We argue that search intermediaries can affect purchase decisions not only by highly ranking products that they would like purchased, but also by highlighting bad products in regions of the attribute space that they would like to push the consumer away from.

*The authors would like to thank Bart Bronnenberg, Jun Kim, and Carl Mela for sharing their search data. We would also like to thank Vasilis Syrgkanis and Eliot Abrams for discussions that inspired this paper, and Liran Einav, Matthew Gentzkow, Tim Bresnahan, Marc Rysman, and participants at IIOC 2017 for helpful comments.

1 Introduction

live in a world where the supply of information is substantial and increasing; it is more widely-shared (through the internet) and catalogued (by search engines). As Herbert Simon presciently argued (Simon 1971), this has substantially increased the time spent acquiring information: searching is now a significant part of the day for many people, and one that we need to understand.[1]

In classic models of sequential search, agents want to choose one item from a set of heterogeneous objects that appear identical prior to search (McCall 1970, Rothschild 1974). Searching an item allows the searcher to learn the payoff to the sampled option. Depending on whether the payoff distribution is known or itself being learned, this results in an optimal stopping problem with a constant or time-varying reservation value.

In reality, in most search contexts some differences between options are ex-ante observable. For example, in online search, search engines rank some options higher than others, and these rankings are observable prior to search. In job search, even before interviewing, job seekers see many observable characteristics of potential options, such as firm, pay-range, job description etc. It is the ex-ante unobservable characteristics that are learned through search. Given that there are observable differences between options, the decision problem is not only when to stop searching, but *which options to search*.

In the special case when the payoffs to each option are conditionally independent (where the conditioning is on the observables), (Weitzman 1979) provides an index-rule characterization of the solution to the consumer search problem. But this model rules out the possibility that consumers learn about the payoff from one option from sampling another.

Our paper introduces the idea of spatial learning: when a searcher samples an option and observes an unexpectedly high or low payoff from that option, they update on the payoffs to other options *that are close in the space of observables.* For example, a job seeker receiving a very attractive offer at Microsoft might reasonably infer that an (unsampled) Google offer would be better than they

[1]For example, (Boik, Greenstein and Prince 2016) show that the average US household participating in the Comscore survey spent around 2 hours a day online in 2013 (although the authors note that some of this is content consumption).

had previously expected, but not update on the value of an offer from McKinsey, say. A student deciding which colleges to apply to may cancel their campus visits to liberal arts colleges after a bad experience with one of them. Or a consumer looking for a camera who reads the reviews for a particular camera with low resolution and decides it is not for her will probably negatively update on all low resolution cameras.

As we show in a simulated example, this learning process leads to *path dependence* in search — a consumer who has a bad initial experience when sampling some part of the product space will tend to focus their search elsewhere in the future. This offers strategic possibilities for a search intermediary: by highlighting worse-than-expected products in some parts of the product space, they can steer consumers away from those areas and towards a desired purchase.

offer a framework for modeling spatial learning. The building blocks are a *characteristic space* consisting of observable salient characteristics of the options (search rank, price, star rating); and utility functions modeled as a *Gaussian process* over that characteristic space. Gaussian processes have found wide application in machine learning (Rasmussen and Williams 2005). They are very flexible and yet are fully specified by a mean function (giving the expected payoff to any unsearched option) and a kernel function (giving the covariance between pairs of options). The kernel function takes as inputs the locations of any two options in characteristic space, and outputs a covariance between them. Searchers will update more about close-by options than far-away options. The kernel specifies the distance metric, and encodes the mental model that searchers use to extrapolate.

Initially, it is optimal to "explore" the characteristic space, learning a model of the payoff of different regions. As search progresses, however, searchers will tend to "exploit" their knowledge, concentrating search in the neighborhood of previously explored high-payoff options. This leads to a "zooming in" of searches around the product eventually purchased, a phenomenon first observed in a study of online digital camera search by (Bronnenberg, Kim and Mela 2016) (BKM).

apply this framework in re-analyzing the BKM dataset. Replicating the results of that paper, we show that the products searched by consumers converge in attribute space to the product ultimately purchased and that the variance of product characteristics searched declines as search progresses. This is consistent with our model's predictions. To these we add the new findings that the difference in characteristics between successive searches declines over the search path (i.e. that search becomes increasingly local in characteristic space), and that consumers tend to take a large step away (in characteristic space) after viewing often searched but rarely purchased products. The latter offers an indirect test of the spatial learning hypothesis: such products presumably have lower payoffs than expected (since they are searched but not bought), and under spatial learning this bad news should depress the expected payoff to all local products, whereas under the null of independence (no learning), this should have no effect.

also fit a structural spatial search model to the data. incorporate random coefficients on mean utility, so that the model nests the often used random coefficients discrete choice model of (Berry, Levinsohn and Pakes 1995) in the special case of zero search costs. The model has the attractive feature that the entire search path — each option considered, as well as the option finally chosen, as well as the order of search — is informative about individual consumer preferences. estimate the model by Markov Chain Monte Carlo under a one-period look ahead assumption on the search process; relaxing this assumption, as well as accounting for the possibility of price endogeneity (which is more subtle here than in the standard complete information model, because of dynamic information acquisition) is left to future work. show that the model is able to rationalize the patterns we observe in the data.

use a counterfactual with the estimated parameters to explore path dependence, by forcing consumers to sample an unrepresentative "bad product" early on in the search process. This misleading information causes some consumers to terminate search early, and others to substitute away, generally reducing welfare. This sheds some light on the market power held by search engines and online platforms, who have considerable power over the order in which products are viewed, which in turn shapes consumption (Dinerstein, Einav, Levin and Sundaresan 2014).

Related literature. Search is a well-studied topic in the microeconomic theory, and the classic models cited earlier have been used as inputs into explaining various economic phenomena, such as price dispersion (see e.g. (Varian 1980)). Recently, (Adam 2001) analyzes a model which allows for payoffs to be sampled from a discrete set of unknown distributions ("nests"), so that searchers who sample an option from one nest will update their posterior on the distribution for all other items on this nest. (Dickstein 2014) applies this sort of model in an empirical study of physician learning in the anti-depressant market. Our contribution to this literature is to allow for a spatial correlation in beliefs, which is more flexible than the models above (though to keep things tractable we impose parametric structure through Gaussian processes).

There has been empirical IO work on the identification and estimation of some of the classic search models (Koulayev 2014), as well work on testing the alternative of sequential versus non-sequential search using Comscore data (De Los Santos, Hortaçsu and Wildenbeest 2012). In the marketing literature, in addition to BKM, there have been a number of papers that have taken the (Weitzman 1979) model to data, including (Kim, Albuquerque and Bronnenberg 2010), (Honka and Chintagunta 2017) and (Ursu 2017).

Paper outline. The remainder of the paper proceeds as follows. Section 2 provides an illustrative example of path dependence in search and spatial learning. Section 3 outlines a general model and derives implications for consumer search behavior. Section 4 describes the data on consumer search paths we use to test our model, and presents stylized facts from this data that match model predictions. Section 5 describes the estimation of the model using data on search paths. Section 6 presents the results of the estimation, and section 7 concludes.

2 An Illustrative Example

We begin the paper with an example that illustrates the main forces present in our model. Consider a world with 3 products, A, B and C. A consumer has to buy one of the three (we add an outside option in the main model, but omit it here for simplicity). Their payoff from consumption depends on price and quality according to:

$$u_j = q_j - p_j$$

Quality is unknown to the consumer ex-ante; all they observe are the prices, which are ordered as $p_A = 2 < p_B = 3 < p_C = 4$. By searching a product, they learn the payoff u_j. Each search costs

$c = 0.4$, and products must be searched before purchase.

Assume that consumers know that $\boldsymbol{q} \sim N(\boldsymbol{p}\mu, \Sigma)$ where $\boldsymbol{q} = (q_A, q_B, q_C)$, $\boldsymbol{p} = (p_A, p_B, p_C)$. $\mu = 1.3$ is a known scalar. Because μ is positive, price acts as a signal of quality, and because it is greater than one, consumers believe that increasing price implies higher expected utility. The variance-covariance matrix Σ is also known ex-ante. Consistent with the spatial logic offered in the introduction, we assume that it takes the form $\Sigma_{ij} = 1.4 \exp(\frac{-(p_i - p_j)^2}{\rho})$, with diagonal elements therefore equal to 1.4.

As an initial baseline, consider a model where $\rho \approx 0$, so that all the off-diagonal elements of Σ are zero and there is no spatial correlation in payoffs. This is a special case of the (Weitzman 1979) model. The optimal search algorithm assigns each option a score z_j — which in our example satisfies $z_A < z_B < z_C$ — and requires searching those in decreasing order of score, stopping if the maximum payoff found thus far exceeds the search index of the next option to be searched.

The solution is illustrated graphically in the left panel of Figure 1. After searching product C, consumers will stop if the observed value of u_C is above the reservation utility z_B, and otherwise will search product B. If the observed utilities u_C and u_B are both below z_A, then the consumer will then search product A. Notice that there is no path dependence; regardless of the utility realizations, consumers will search products in the order C, B, A.

Next consider a model in which $\rho = 0.8$, so that payoffs are spatially positively correlated. Since $|p_A - p_C| > |p_A - p_B| = |p_B - p_C|$, consumers will update more about B than the remaining products after exploring either A or C. There is no straightforward characterization of the optimal search strategy anymore, and we solve for it numerically by backward induction. The right panel of Figure 1 illustrates the results of this exercise. As before, the consumer starts by searching product C. But the next product they search depends on the observed value of u_C. If u_C is sufficiently high (the yellow region), they stop and buy it. If u_C is intermediate, they move on to product B, buying either B or C if B is good enough (brown region), and only searching A if the max of B and C is low (blue region).[2] If u_C is low, they infer that μ is also low, and instead target product

[2]The values of u_B and u_C matter individually too. The downward sloping line at the top of the blue region indicates that for a fixed u_B just above 0.8, the decision to search A depends on whether the news about C was

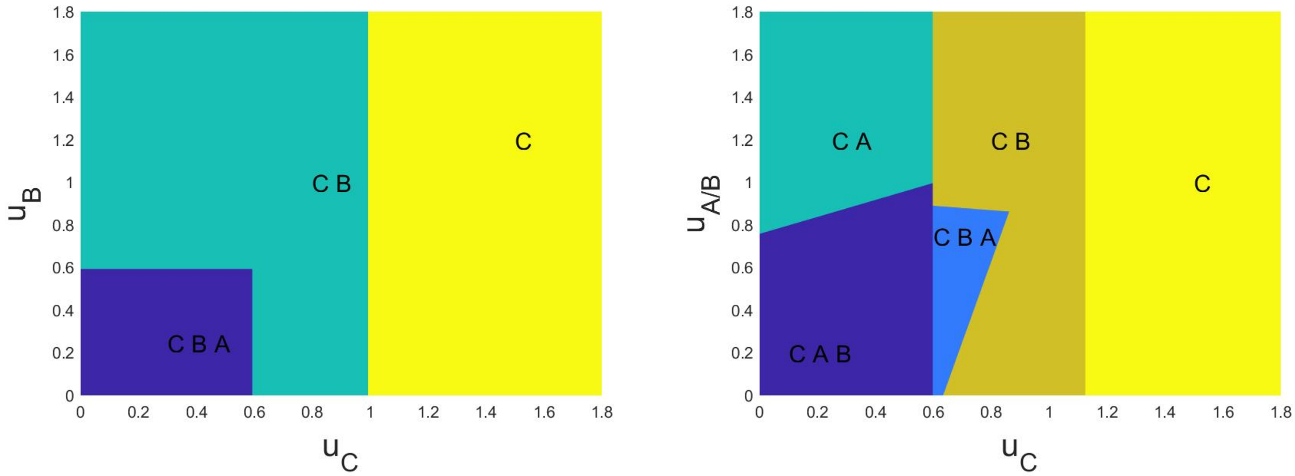


Figure 1: **Optimal Search Strategies.** The left panel shows the optimal search strategies when there is no correlation in quality across products, and observing u_j only provides information about product j. The right panel illustrates how search strategies change when consumers believe that there is positive cross-product covariance in quality. The x-axis is the realized utility of the first product searched, and the y-axis is the realized utility of the second product searched. Each region records the order in which products are searched before the consumer stops searching. In this example, $p_A = 2$, $p_B = 3$, and $p_C = 4$. $\mu = 1.3$, $c = 0.4$, $\Sigma_{ii} = 1.4$, and $\Sigma_{ij} = 1.4\exp(\frac{-(p_i-p_j)^2}{\rho})$. In the left panel, $\rho \approx 0$ and in the right panel $\rho = 0.8$.

A next, moving onto product B (purple region) if the maximum payoff of A and C is sufficiently low, and otherwise stopping (green region). So there is path dependence: each successive outcome determines whether to stop and where to go next.

Search Rankings and Manipulation of Beliefs. In our example so far, we have assumed equal costs of searching all products. But it is well documented (for example, see (Ursu 2017)) that the ranking or salience of products on online platforms affects the order in which consumers search through those products. We thus allow for different search costs, where higher-ranked products have lower search costs. The direct effect of this is to ensure that higher-ranked products are more attractive to search. But under spatial learning, there are also spillover effects: what consumers learn from searching a highly-ranked product can affect consumers' beliefs about other products. *Product rankings can therefore be used to manipulate both search costs and beliefs.*

To show the ways in which rankings can be used to change purchase behavior, we modify our

good. If it was good, then the posterior μ is higher and price is a stronger signal of quality, so it is optimal not to search A; whereas if it was bad the converse applies.

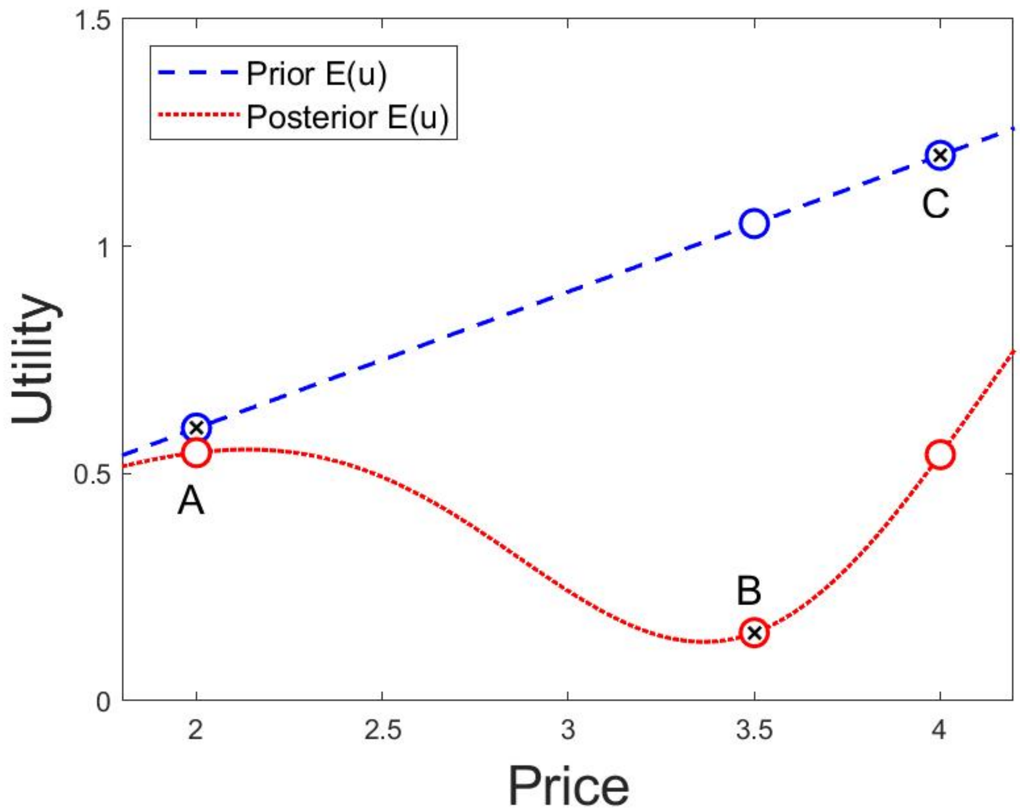


Figure 2: **Belief updating.** Black crosses indicate the location of the three products, A, B,and C, in price-utility space. The blue dashed line is the consumer's prior expected utility of hypothetical products at different price levels. This is given by $E(u_j) = (\mu - 1)p_j$. The red dashed line indicates the consumer's posterior expected utility at different price levels after searching product B. The posterior is computed using the bayesian updating rule described in the text. Parameters are as described in the notes to Figure 1.

example from before by setting $p_B = 3.5$ so that product B is closer in price space to product C than product A. We set the search cost for product B to zero, so that searching it is free — and therefore it is optimal to search B first (this is an attempt to model it being heavily promoted by the platform). Last we assume that the latent payoff for B is $u_B \approx 0.2$, much worse that expected.

Figure 2 illustrates how the consumer's beliefs about u_A and u_C are updated after she searches product B. This bad initial experience drags down the posterior mean beliefs about C more than product A, so that after the "free" search of B, the consumer believes that A is a better option.

This "belief manipulation" can be effective in driving consumers towards a desired option. Suppose for example that the search intermediary wants the consumer to buy product A, perhaps because it earns the highest commission on sales from that seller or because it is a "house brand". Intuitively, one might expect that the best the intermediary can do is to promote product A, driving its search cost to zero and ensuring it enters the consumer's consideration set. Yet it turns out the answer is more subtle and depends on the search costs.

Table 1 records the consumer's purchased product as a function of the product they are shown first, and the search cost, c. For low search costs ($c < 0.05$), the consumer will search every product and ultimately purchase C, the highest utility product. In this search cost regime, the platform cannot control the purchase outcome. On the other hand, for very high search costs ($c > 0.91$), the consumer will not search beyond the product initially shown to them by the platform. The platform has complete control over the purchase decision, and therefore should show product A first so that it is purchased. The surprise is that in intermediate cases ($c \in (0.05, 0.78]$), the platform can achieve its aim of getting product A purchased only by showing *product* B first. If the consumer views either A or C first, the observed utility will be equal to the prior expected utility, and the consumer will not update their expectation about the other products. Thus, if the consumer is shown product A first, she will search product C second, since $E(u_C|u_A) = (\mu - 1)p_C > (\mu - 1)p_B = E(u_B|u_A)$. After viewing C she will purchase it. However, if she is shown the inferior product B first she will infer that product C is likely to be low quality since it is close to product B in price space, and will therefore search product A second. With intermediate search costs, it is then optimal to stop and purchase product A.

Notice that this "intermediate case" is likely to be the most prevalent in practice, since we think of platforms as having some but not perfect control over what is purchased on their sites. They also often have considerable prior data on purchasing decisions which may allow them to predict with high accuracy which products are "surprisingly bad" and may therefore be used to steer consumers in this way (we ourselves do such prediction using a relatively small Comscore dataset later in the paper). So belief manipulation is a realistic possibility, depending on the motivation and sophistication of the search intermediary.

Starting Product	c∈ $[0, 0.05]$	c∈ $(0.05, 0.78]$	c∈ $(0.78, 0.91]$	c∈ $(0.91, \infty]$
A	C	C	C	A
B	C	A	B	B
C	C	C	C	C

Table 1: **Purchase as a Function of Starting Product and Search Cost.** Each cell records the product purchased by a consumer with search cost c given by the column headers who is shown the starting product indicated by the first column before starting to search. Parameters are as described in the notes to Figure 1.

3 Model

In this section, we develop a general model of consumer search with spatial learning. Consumers face a set of products with known attributes, but are uncertain about the mapping of product attributes to payoffs. Searching a product reveals its payoff. Since the consumer believes that utility is correlated across products that are close in attribute space, searching one product leads the consumer to update her beliefs about the payoffs to other options.

3.1 Environment

A consumer with unit demand faces a finite set $\boldsymbol{J}$ of available products. Each product has a set of characteristics $X_j \in \boldsymbol{X} \subseteq \mathbb{R}^K$ that are observable to consumers before search. Each product also has an associated search cost c_j. By paying the search cost, the consumer may learn the payoff u_j from buying product j. Consumers may search as many products at they like. After terminating search, they may consume any product they have searched (they may not purchase a product without searching it first) or choose to consume the outside option instead, with payoff $u_0 = 0$. Their final utility is the payoff from the product consumed, less the sum of the search costs.

We assume that the payoffs have the following structure:

$$u_j = m(X_j) + \epsilon_j \tag{1}$$

where $m : \boldsymbol{X} \to \mathbb{R}$ is a function that maps a vector of characteristics to average payoffs, and ϵ_j is an idiosyncratic shock sampled iid across products from a distribution $N(0, \sigma)$. The function $m(X)$ is unknown to consumers, and sampled from a Gaussian process with prior mean function

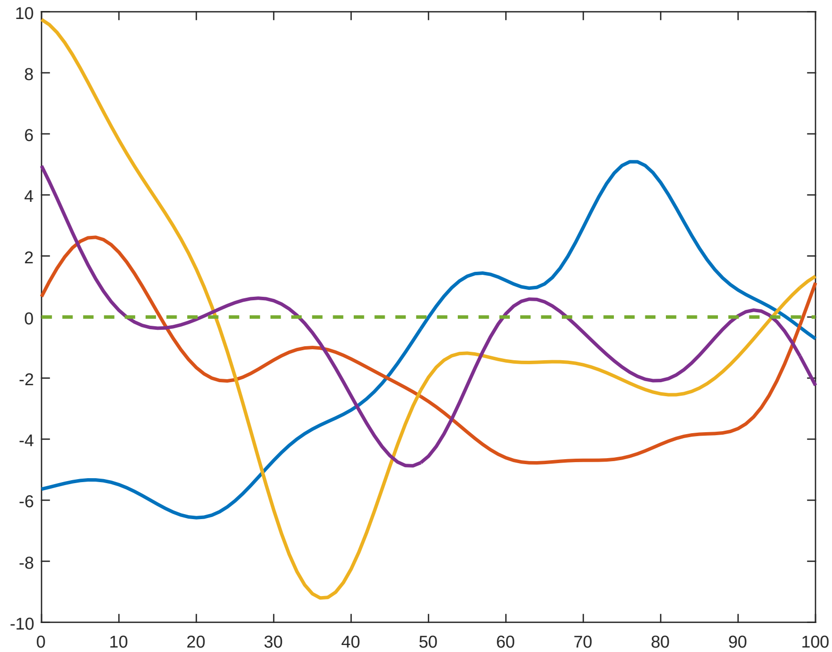


Figure 3: **Gaussian Process Draws** This figure illustrates draws from a single dimensional Gaussian process with mean 0. The dashed line is the mean of the process, and the solid lines are draws from the process. The x-axis is the single dimension over which the process is defined.

$\mu(X)$ and covariance function $\kappa(X, X')$. We assume that μ is a continuous function, and that $\kappa(X, X') \equiv \tilde{\kappa}\left(\frac{\|X-X'\|}{\rho}\right)$ for some weakly positive, continuous and decreasing function $\tilde{\kappa}$, where $\|\cdot\|$ is the Euclidean norm and ρ is a bandwidth parameter. We assume that the consumers know the prior. As consumers search, they update their beliefs about $m(X)$ according to Bayes rule (see the section on beliefs and learning below). Figure 3 illustrates four draws of a Gaussian process with $\mu(X) = 0$ and the square exponential covariance function $\lambda \exp\left(-\frac{1}{2}\left(\frac{\|X-X'\|}{\rho}\right)^2\right)$ with $\lambda = 10$, $\rho = 10$, and $X \in [0, 100]$.

An interpretation of the model is that initially consumers don't know their preferences over characteristic space, which are summarized by $m(X)$. As they search, they get noisy signals of the function (noisy because of the ϵ shocks). Because $\tilde{\kappa}$ is decreasing, the covariance in payoffs declines with distance in characteristic space, and learning is spatial: sampled payoffs are more informative about the payoffs of close-by products than those far away.

Another interpretation consistent with the model is that consumers know their preferences over the observable characteristics X, but there are other unobservable product characteristics whose values are unknown without search. As they search, consumers refine their model of the mapping between the observable and the unobservable characteristics, updating the model $m(X)$.

Two special cases are worth noting. As $\rho \to 0$, the correlation in average payoffs between any two points goes to zero, so that each product has independent and unknown payoffs prior to search. This is the model of Weitzman (1979), specialized to the case of normally distributed payoffs. As $\rho \to \infty$, the correlation in average payoffs goes to one, so that learning the payoffs at any one point is equally informative for all other points.

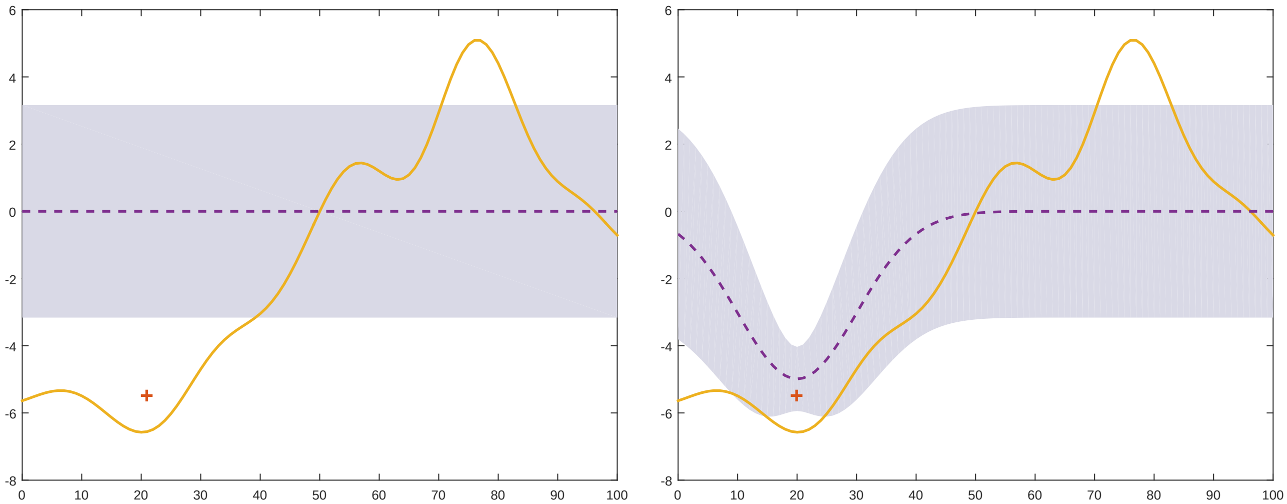
3.2 Beliefs and Learning

The search process is a non-stationary Markov Decision process. We choose to model the state as a tuple $S_t = (\mu_t(X), \kappa_t(X, X'), \hat{j}, \hat{u}, \mathcal{J})$, where $\mu_t(X)$ are the current mean beliefs, $\kappa_t(X, X')$ is the current covariance, $\widehat{j}$ is the best product found so far, $\widehat{u}$ is the payoff to the best product found so far and $\mathcal{J}$ are the available products remaining to be searched. The transitions on the state variables $\hat{j}, \hat{u}, \mathcal{J}$ are straightforward. The mean and covariance functions update according to:

$$\mu'(X) = \mu(X) + \frac{\kappa(X, X_j)(u_j - \mu(X_j))}{\kappa(X_j, X_j) + \sigma^2} \tag{2}$$

$$\kappa'(X, X') = \kappa(X, X') - \frac{\kappa(X, X_j)\kappa(X_j, X')}{\kappa(X_j, X_j) + \sigma^2} \tag{3}$$

Figure 4 illustrates the consumer's learning process. Panel A represents a consumer's prior beliefs and ex-ante unknown preferences over a one-dimensional characteristic space $X \in [0, 100]$. The consumer's prior mean, $\mu(X) = 0$ is indicated by a dashed line. The shaded area is a one standard deviation band of the prior Gaussian process around the mean. The solid line is the consumer's utility function $m(X)$ which is drawn from the Gaussian process. The consumer searches a product j and observes the utility u_j, indicated by the by the point in Panel A. Panel B shows the consumer's posterior beliefs. Notice that the observation has reduced the consumer's uncertainty about her utility function $m(X)$, especially for products close to X_j in parameter space.



(A) Prior Beliefs (B) Posterior Beliefs

Figure 4: **Gaussian Process Learning** This figure illustrates Bayesian updating in a single dimensional Gaussian process with mean 0. In Panel A, the dashed line is the prior mean, and the shaded area is a one standard deviation interval around the mean. The solid line is the "true" function which is drawn from the Gaussian process, and the cross is the value observed by an agent, which is equal to the value of the Gaussian process draw plus noise. In Panel B, the dashed line reflects the mean of the agent's posterior beliefs. The shaded area is a one standard deviation interval of the posterior beliefs.

3.3 Consumer Behavior

Because there are a finite set of products that can be searched, the consumer's decision problem can be solved by backward induction. Doing so will generally be computationally intractable with a reasonable number of products, since the state variables are continuous functions.[3]

We would like to get a closed form solution so that we can illustrate some of the forces in the model. Accordingly, we analyze a heuristic solution to the problem: one period look ahead search. This is a common solution method in the literature on Gaussian processes, which typically employ n-period look ahead assumptions (Osborne, Garnett and Roberts 2009).

The one-period look ahead policy is a greedy policy: it scores the available options based on their expected marginal contribution over the current best option $\hat{u}$. Define $s_j = \sqrt{\kappa(X_j, X_j) + \sigma^2}$, the standard deviation of the payoff of product j (which includes the idiosyncratic shock). Define $a_j = (\hat{u} - \mu(x_j))/s_j$, the current best option normalized by the mean and variance in payoffs for

[3]One could instead take the set of realized utilities for searched objects as the state variable — since these are sufficient for the beliefs — but with a large number of products this remains intractable. For example, Crawford and Shum (2005) interpolate the value function between a discrete set of states in a setting with 5 products. In section 5 we apply the model to a setting with around 300 products.

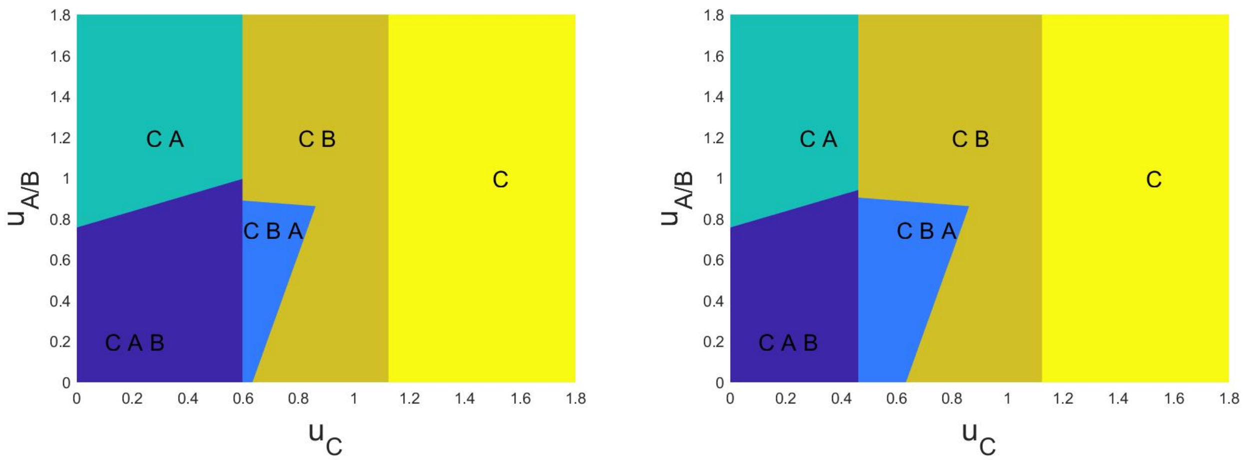


Figure 5: **Policy approximation.** The left panel shows the optimal policy and the right panel the optimal one-period look ahead policy, in the example from earlier, with $p_A = 2$, $p_B = 3$, and $p_C = 4$. $\mu = 1.5$ $c = 0.45$, and $\Sigma_{ii} = 0.55$, $\Sigma_{AB} = \Sigma_{BC} = 0.15$, and $\Sigma_{AB} = 0.005$.

item j. Then we score option j according to:

$$z_j = \Phi(a_j)\hat{u} + (1 - \Phi(a_j))\mu_j + \phi(a_j)s_j - c_j \tag{4}$$

where the first term captures the chance that product j is not better than the current best, the second two are the expected value of product j conditional on being better times the probability of that event, and the last term subtracts the product-specific search cost. The one-period look ahead policy is to search the option with the highest score z_j, so long as it exceeds $\hat{u}$; otherwise to stop and buy the current best option. While one-period look ahead is not generally optimal, it can often provide a close approximation to the optimal policy. Consider the example of Section 2 above. In that case, the optimal and one-period look ahead policies nearly coincide: Figure 5 shows the globally optimal policy in the left panel and the optimal one-period look ahead policy in the right panel.[4] For our Gaussian process environment, one-period look ahead is also exactly optimal in some special cases.

Proposition 1 (Optimality in Special Cases). *When the mean payoffs are perfectly correlated ($\rho = \infty$) or independent ($\rho = 0$), one-period look ahead is exactly optimal.*

[4] The main difference is that after observing $u_C \in [0.5, 0.6]$, the optimal policy searches A next since this is more informative than searching the middle product B (products A and C together bracket B, providing good signals of its payoff). One period look ahead does not account for the value of information beyond the option searched.

Proof. Under independence, the beliefs never update and consequently the order of search is predetermined. The optimal order of search and stopping rule is given by Weitzman (1979). Since in our environment the payoff distributions are *a priori* identical up to mean shifts, the optimal order of search is in descending order of prior mean utilities, stopping when the marginal gain from search is less than the search cost. The one-period search policy ranks the options in the same order — by prior mean utility — and has the same stopping rule, and so the two policies coincide.

Under perfect correlation, the mean beliefs update everywhere symmetrically so that each update $\mu' - \mu$ is constant in X. The updates thus don't change the underlying decision problem — they are just affine transformations of the utility — and thus the problem is essentially iid. The optimal search order is fixed and still follows Weitzman (1979), while the stopping rule should consider searching one more time and then stopping versus stopping now. As noted above, the one-period look ahead algorithm implements the Weitzman algorithm in our setting, and thus has the same stopping rule. □

We can also prove some comparative static properties, using the analytical characterization of the score in (4).

Proposition 2 (Comparative statics)**.**

$$\frac{\partial z_j}{\partial \hat{u}} = \Phi(a_j) > 0 \ , \ \frac{\partial z_j}{\partial \mu_j} = 1 - \Phi(a_j) > 0 \ , \ \frac{\partial z_j}{\partial c_j} = -1 < 0 \ ,$$

Moreover the impact of the payoff to the last search u_k on current scores is given by:

$$\frac{\partial z_j}{\partial u_k} = \frac{\partial z_j}{\partial \mu_j}\frac{\partial \mu_j}{\partial u_k} + 1(u_k = \hat{u})\frac{\partial z_j}{\partial \hat{u}} = (1 - \Phi(a_j))(\kappa(X_j, X_k)/s_k^2) + 1(u_k = \hat{u})\Phi(a_j)$$

Proof.

$$\begin{aligned}\frac{\partial z_j}{\partial \hat{u}} &= (\phi(a_j)/s_j)\hat{u} + \Phi(a_j) - (\phi(a_j)/s_j)\mu_j + \phi'(a_j)s_j/s_j \\ &= \phi(a_j)(\hat{u} - \mu_j)/s_j + \Phi(a_j) + \phi'(a_j) \\ &= \phi(a_j)a_j + \Phi(a_j) - a_j\phi(a_j) \\ &= \Phi(a_j)\end{aligned}$$

where on the third line we use the fact that $\phi'(x) = -x\phi(x)$. Similarly:

$$\begin{aligned}\frac{\partial z_j}{\partial \mu_j} &= -(\phi(a_j)/s_j)\hat{u} + (1 - \Phi(a_j)) + (\phi(a_j)/s_j)\mu_j - \phi'(a_j) \\ &= 1 - \Phi(a_j)\end{aligned}$$

The partial on costs is immediate. Finally, notice that from the transition equations (2) and (3) the last observation's payoff only influences mean beliefs and potentially the current highest payoff $\hat{u}$ (if it was better than the prior best option). Applying the chain rule, we can use the partial derivatives derived above, along with the derivative $\frac{\partial \mu_j}{\partial u_k}$ from (2) to get the result. $\square$

These properties are intuitive, but have some interesting implications. First, an improved current best option affects the score of a product at a rate that depends on whether its payoff may fall below the best option - i.e. based on the tail risk of an option. It follows that consumers score risky options more highly when they have better existing options. Second, the comparative static on search cost implies an important role for product rankings and visibility in driving search paths. Finally, the main beneficiaries of a higher payoff for the last search are options that have high covariance with the last search location (close by products, provided the area is not already well explored), with good potential of beating the current best option (since otherwise the higher mean payoff for that product does not translate into a higher expected utility from search). In the event that the last option was the best so far, there is an additional term that favors riskier options.

3.4 Path Dependence and Platform Power

One of the main features of the model is path dependence. This follows immediately from Proposition 2 above, which shows that the outcomes of past searches affect current search scores, which in turn impacts the order of search. This may be contrasted with standard sequential search, where the products are not ex-ante differentiated, and also with the Weitzman (1979) model where the order of search is fixed in advance and does not depend on experiences along the way.

Path dependence allows search intermediaries the ability to influence purchase behavior in subtle ways. It is well known that search intermediaries can direct consumer attention to particular products by ranking them highly in search results. As Ursu (2017) has shown, using an experiment ran by Expedia, highly ranked items are clicked more frequently even if the search results are random. In our model, this ability to influence what consumers click translates directly into an ability to influence beliefs by placing unrepresentative items at the top of the rankings. For example, placing an item with a payoff $u_j >> \mu(X_j)$ at the top of the rankings will increase its probability of search (because c_j is small, implying z_j higher) and have a subsequent "halo effect" on the items with high covariance with j, which will generally be the items with small distance from X_j. Conversely, placing a surprisingly bad item — one with $u_j << \mu(X_j)$ — at the top of the rankings will tend to drive search towards a different part of the product space, increasing the likelihood of purchase in that part of the space. As we illustrated in the example in Section 2 above, a platform that wants to promote purchase in one part of the space ("the focal region") may optimally "drag down" beliefs about competing regions rather than promoting the focal region. To do so, the platform must be able to manipulate search costs (by re-ranking) and also have knowledge (or good signals) of the payoff to the various products, so that it knows which products are surprisingly good or bad. An estimation procedure based on historical data, such as the one we use in the empirical section below, can be used to do this.

	Mean	SD	Min	Max
Price	302.13	478.47	16.99	5250.00
Zoom	6.04	5.60	0	35
Pixel	10.54	3.13	1	21
Display	2.67	0.41	1.1	3.5
N	357			

Table 2: **Summary Statistics: Products** Products from the digital camera data are defined by unique values of brand, zoom, pixel, and display. If there are multiple prices recorded for the same product, we use the average price recorded over all searches.

4 Empirical Evidence from Online Search

4.1 Data

We apply our model of consumer search with spatial learning to data which records the search paths of consumers shopping online for digital cameras. The data comes from ComScore, who track the online browsing behavior of panelists who have installed ComScore's tracking software. The sample we use was constructed by Bronnenberg, Kim, and Mela (2016) (henceforth BKM), and comprises the browsing activity of 967 ComScore panelists who were searching for digital cameras between August and December 2010.

For an individual panelist, we observe the sequence of products viewed, the product eventually purchased, and the date and time of each observation. Product views were detected by scraping the sequence of URLs visited by consumers for product information. The data covers all browsing behavior and therefore is not limited to one retailer. A product "view" or "search" (we use the terms interchangeably) in the data is recorded when a webpage providing information about a single product is loaded. This could include product pages on retail sites such as Amazon.com, manufacturer websites, and review sites. Purchases are identified using a second ComScore dataset that tracks online transactions carried out by panelists. For each product view, the data records the product make and model, four continuous product attributes - price, zoom, display size, and

		Mean	Median	SD	Min	Max
	Search Length	5.59	4.00	6.52	1.00	58
	Chosen Product Discovered	0.79	1.00	0.29	0.03	1.00
	N	966				
Price	Purchased	211.46	129.00	264.85	16.99	2862.03
	Searched	285.45	150.00	406.45	16.99	5250.00
Zoom	Purchased	6.42	4.00	6.02	0.00	35.00
	Searched	6.43	4.00	5.97	0.00	35.00
Pixel	Purchased	12.42	12.20	2.16	1.30	21.00
	Searched	11.96	12.10	2.37	1.00	21.00
Display	Purchased	2.82	2.70	0.26	1.10	3.50
	Searched	2.78	2.70	0.30	1.10	3.50

Table 3: **Summary Statistics: Search Paths** Table summarizes all search paths in the digital camera data. Search path length is the number of unique products viewed. The second row records the search percentile, as defined in the data, at which the product eventually chosen is first viewed. The lower panel records the distribution of attributes over searches and purchases. For example, if a certain product is searched by several consumers but never purchased, it will enter the distribution of searches once for each consumer that searched it, bu tit will not enter the distribution of purchases.

pixels - and a number of binary attributes including whether the camera is an SLR, whether the camera has video capability etc. The conversion of the raw ComScore browsing data and the matching of this data to product attributes was performed by BKM, and extensive details on the preparation of the data are provided in that paper. Note that this is a selected sample and not representative of the population of consumers. We use this data to illustrate broad patterns that motivate our modeling approach and to test our model.

BKM allow the price of a product to vary over time and across website domains. To simplify our analysis, we define a product as a unique combination of brand, pixel, zoom, and display. We then take the average price recorded for that combination in the data to be the price of that product.This leaves us with 357 products described by four continuous attributes (price, zoom, display size, and pixels), as well as a discrete characteristic, camera brand. Table 2 records summary statistics on the distribution of these attributes across products.

An observation in the data is a sequence of products viewed and the identity of the product purchased. We drop repeat searches from the data, keeping only the first view of a product in each consumer's search path.[5] Table 3 records summary statistics on consumer search paths. The first row of the table records path length - the number of products searched before purchase. Note that the product which is purchased is always one of the products searched. The average consumer views about 5.6 products. There is a tail of consumers with very long search paths, the longest of which is 58 products. The second row documents the search percentile at which the ultimately purchased product is first discovered. If a consumer searches N products in total, then the search percentile of the nth product is $\frac{n}{N}$. Note that the Nth product is not necessarily the product purchased. The chosen product is typically discovered towards the end of search. The lower panel of Table 3 documents the distribution of attributes among products *searched* and *purchased*. For example, the mean price of products searched in \$285.45. This is the average over all searches by all consumers - including multiple counts of the same product if multiple consumers search that product. The distribution over products purchased in defined analogously. Notice that the standard deviation of price for products searched is larger than for products purchased, and the

[5] In our model, as well as the rest of the search literature, consumers become perfectly informed after the first time they view a product, so subsequent views are hard to rationalize.

average product searched is significantly more expensive than the average product purchased. This suggests that more expensive products are rarely purchased conditional on being searched. For the other product attributes, the distributions of searched and purchased products are much closer - indicating that those very expensive products that are rarely purchased are not necessarily the products with the highest zoom, display size or pixels.

4.2 Stylized Facts

In this section we present several stylized facts about the search path data. In particular, we describe how consumers move through the four dimensional product attribute space as they search. We argue that these descriptive statistics suggest that consumers begin search with some uncertainty about their preferences over these four attributes, and that they update their beliefs about their preferences for un-searched items after viewing each product in their search path.

Figure 6 replicates one of the main findings of BKM - that the attributes of products searched get closer to the attributes of the product eventually purchased as search progresses. The x-axis of each panel is the search percentile. The y-axis of each panel is the absolute difference in terms of the magnitude of a continuous product attribute between the product searched at that search percentile and the product eventually purchased. The four panels record how distance from the purchased product in each of the four continuous attributes changes over the search path. All four panels display a clear pattern: the attributes of the product being viewed get closer to those of the product eventually purchased over time. This result is partially driven by the fact the the purchased product tends to be first discovered towards the end of the search path, but the pattern persists even if only those products that are *not* purchased are included - products considered, but not purchased, in late search are more similar to the purchased product than products considered in early search.

A second finding of BKM is that consumers search a wider variety of products early in the search path than later in the search path. That is, consumers are not only getting closer to the purchased product in attribute space, but are focusing on smaller areas of the attribute space as search

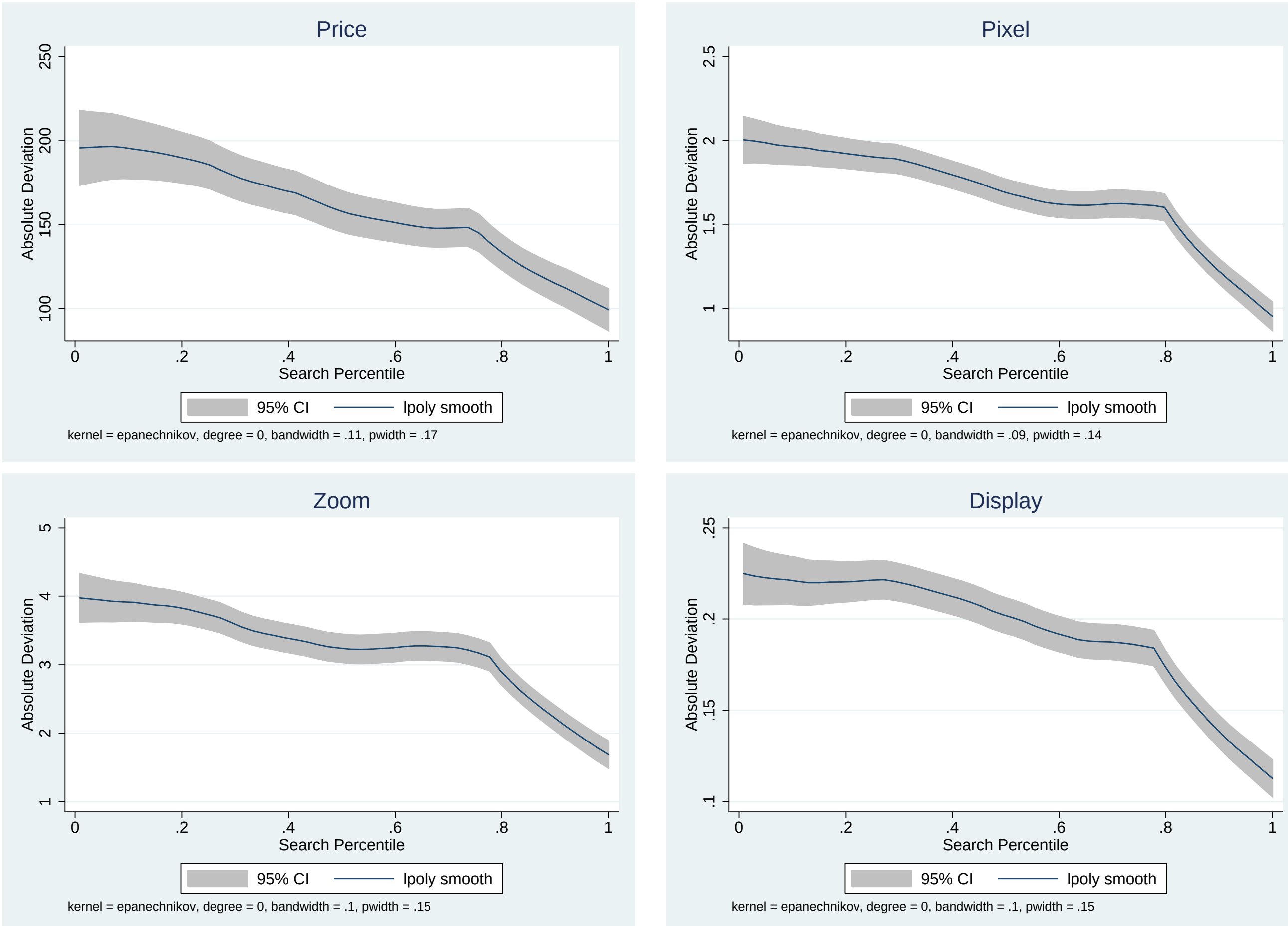


Figure 6: **Convergence to Chosen Attribute Levels** The y-axis for each panel records, for the relevant product attribute, the absolute difference between the searched product and the product ultimately purchased. The x-axis reports the search percentile, as defined in the text. The solid line is a kernel regression using an Epanechnikov kernel, and the shaded area is 95% confidence interval. The estimation sample includes all search paths from the ComScore data on search for digital cameras.

progresses. Figure 7, which also replicates a figure from BKM, illustrates this narrowing of search by plotting the distribution of prices searched in each decile of the search path, where the nth search of a search path of length N is in search decile d if $\frac{d-1}{N} < \frac{n}{N} \leq \frac{d}{N}$. Prices are normalized by taking the difference in log price from the price of the product eventually purchased. The figure shows that the distributions of prices searched in the first search deciles are more spread out than in later deciles. For example, the interquartile range in normalized log price is 2.62 for the 1st decile and 1.83 for the 10th decile. Note that this finding of a "narrowing" of search is not necessarily implied by the convergence of search to the chosen attribute levels illustrated by Figure 6: it could be that consumers always search a narrow area of the attribute space, but move their focus towards the chosen product over time.

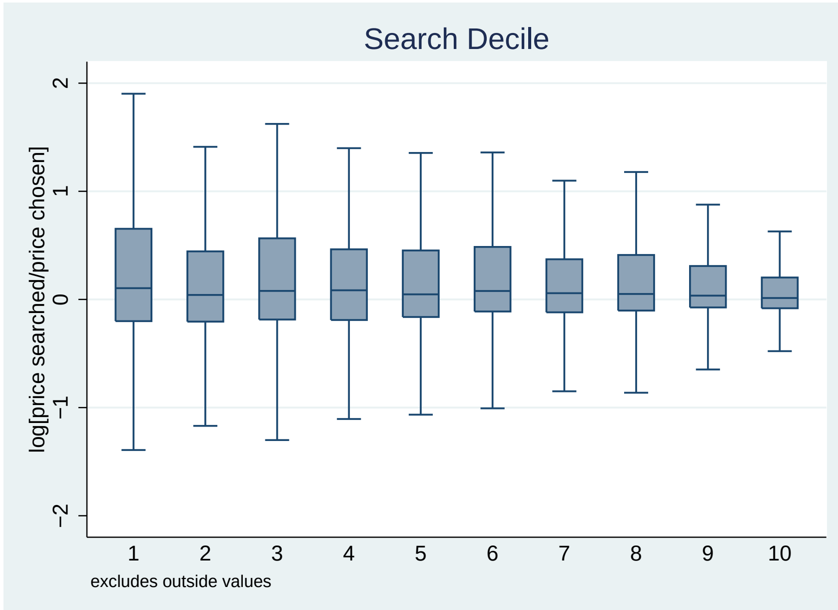


Figure 7: **Convergence to Chosen Attribute Levels** The x-axis records search decile as defined in the paper. Each box plot records the distribution of the log difference in searched price from the price of the product ultimately purchased. The Box records the 25th, 50th, and 75th percentiles of the distribution and the whiskers record the upper and lower adjacent values. The estimation sample includes all search paths from the ComScore data on search for digital cameras.

Figure 8 supports the finding that consumers gradually narrow the scope of their search. For each of the four continuous product attributes, the y-axis records the average "step size" in that attribute, normalized by the average distance between products. For example, a consumer's nth search has a step size in price of $\frac{|price_{n-1} - price_n|}{meanpricediff}$ where $price_n$ is the price of the consumer's nth searched product, and *meanpricediff* is the average absolute difference in prices across all pairs of products. The x-axis records search percentile, as in Figure 6. The results indicate that step size is declining over the search path in all four continuous attributes. For example, in early search the average step size in price is around 60% of the average distance between products, falling to just over 40% by the end of the search path. As with Figure 6, it should not be surprising that step size follows a similar path in all four attributes since product attributes are highly correlated.

Taken together, these patterns suggest that consumers explore a wider variety of products early in their search before narrowing in on close substitutes to the product that is ultimately purchased. This behavior is difficult to rationalize using standard models of search. The most basic such models (see for example McCall, 1970), have consumers drawing utilities at random from a known distribution. These models have nothing to say about the path a consumer takes through attribute space. The fact that there are distinct patterns in the path of search, for example the convergence recorded by Figure 6, would be impossible to rationalize using such models. Others, such as that of Weitzman (1979), allow consumers to have different prior utilities over different objects. For example, in a version of Weitzman's model, utility is composed of a non-stochastic part that depends on product attributes and an i.i.d. stochastic part. Consumers take utility draws from products in an optimal order that depends on product attributes - for example starting with the product with the highest reservation utility and then moving to the second highest etc. Crucially, the order of products searched in such models is fixed, and the stochastic draw observed after each search only influences the decision of whether to continue searching or stop, not the choice of object to be searched next. The findings documented in Figures 7 and 8, that the scope of search narrows, strongly suggests that the information that consumers obtain from early search determines the path of later search. This type of path dependence is hard to rationalize without a model in which there is a spillover of information between searched and un-searched objects.

The model developed in Section 3 allows for exactly this type of information spillover. When consumers view one product they make inferences about their preferences for similar products. Search therefore resolves uncertainty about the utility of un-searched objects. In particular, we assume that beliefs are correlated in attribute space, so that after viewing an object, consumers learn their utility for that object and update their beliefs about other objects - with beliefs about objects that are "nearby" the searched object in product space updated more than objects that are "distant" in product space. Proposition 2 implies that when an object is observed to have a higher than expected utility, other objects that are nearby in attribute space move up the search ranking more than objects that are distant in attribute space. Likewise, when a searched product had lower than expected utility, objects that are closer in attribute space move down the search ranking more than distant objects.

These implications of the model are difficult to test directly, since we do not observe consumer preferences, and hence we do not know what a particular consumer "learns" when he views a

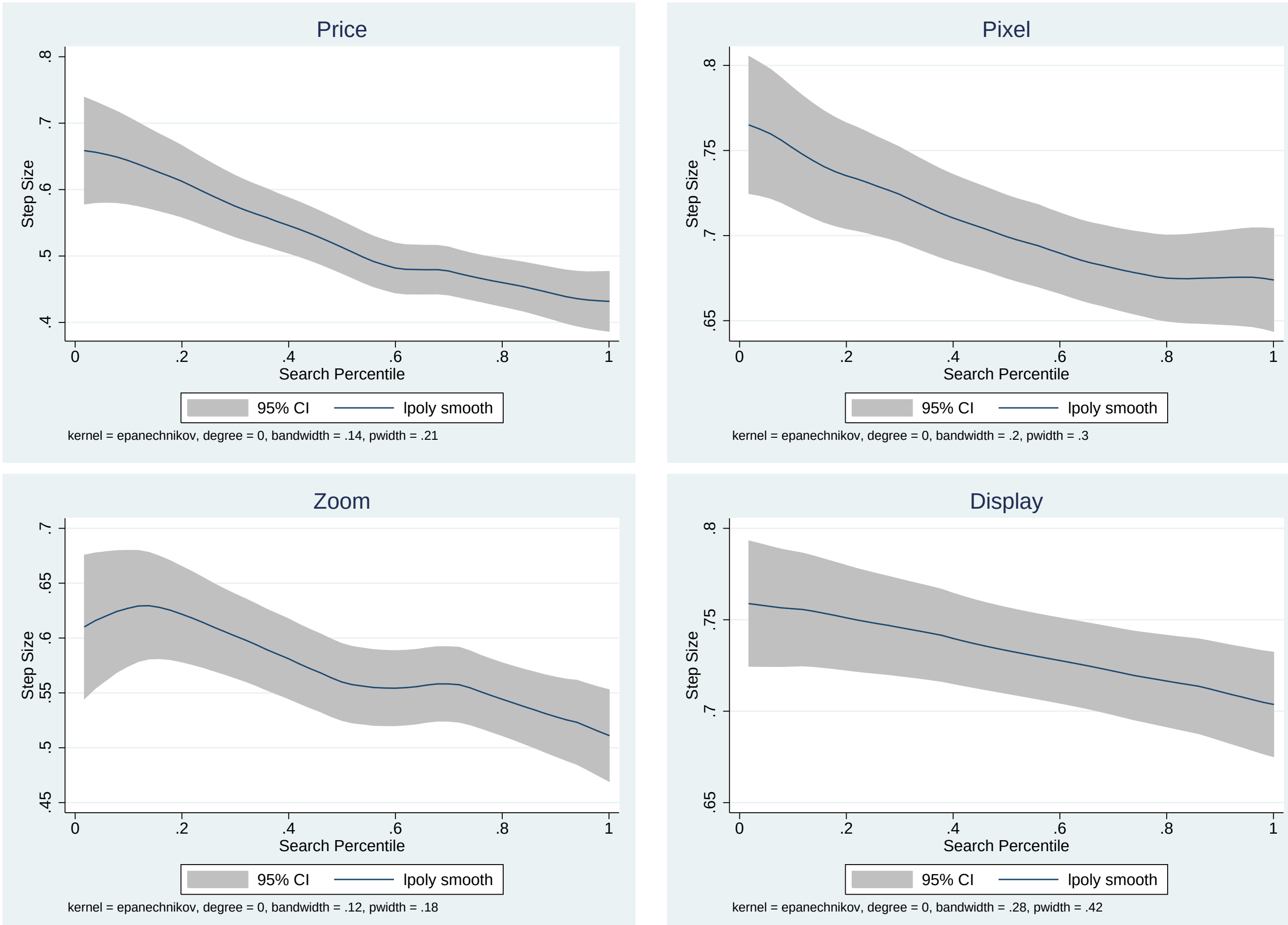


Figure 8: **Step Size** The y-axis for each panel records, for the relevant product attribute, the absolute difference between the product being searched and the product previously searched, normalized by the average difference between products, as detailed in the text. The x-axis reports the search percentile, as defined in the text. The solid line is a kernel regression using an Epanechnikov kernel, and the shaded area is 95% confidence interval. The estimation sample includes all search paths from the ComScore data on search for digital cameras.

particular item, nor what his beliefs are before searching. An ideal experiment would follow the search paths of two consumers with identical prior beliefs about the utility that an object will yield, one with a high idiosyncratic preference for that object and one with a low idiosyncratic preference for that object. After viewing that object, the consumer with a low idiosyncratic preference should make the inference that similar objects also yield low utility, and should be less likely to subsequently search nearby products than the consumer with a high idiosyncratic preference.

The regressions in Table 4 approximate this intuition. The dependent variable of each regression is the absolute step size in a particular product attribute between search t and $t-1$ for individual i - the four panels of the table replicate each regression for the four continuous product attributes. The main explanatory variable, $BadProduct_{it-1}$, is an indicator for whether the last searched product by consumer i was a "surprisingly bad product". These are products that are frequently viewed, but rarely purchased. In this application, we use products that were searched by at least 10 individuals, but were purchased by less than 5% of those who viewed them. The regressions suggest that consumers take larger than average steps in attribute space after viewing one of these products. The first column of the top left panel indicates that the average step size in the price dimension after searching one of these products is $273.2, 58% higher than the step size in price after searching a product not in this category. This pattern is repeated for each of the four product attributes. Controlling for search percentile, as recorded in the second column, does not change this result significantly. The third column controls for the number of products "nearby" the last searched product in attribute space. If "surprisingly bad" products tend to be located in regions of the attribute space that are sparsely populated by other products, then step size after searching one of these products will mechanically be larger. For each product attribute, we add the number of products that lie within one standard deviation (of the distribution of the attribute across products) of the last searched product as a regressor. Four regressors, one for each dimension, are therefore added to each regression. The coefficient on $BadProduct_{it-1}$ in the third column of each panel is closer to zero, but still positive and significant. This suggests that the location of bad products in sparsely populated regions of the attribute space explains part, but not all, of the observed effect.

In the fourth column, we add consumer fixed effects, to account for the possibility that it is heterogeneity in consumers that is driving our results (specifically, that consumers who select "surprisingly bad products" also tend to take larger steps in attribute space). The sign of the coefficient of interest is still positive for all four panels, but only the coefficients in the price and display size regressions are still significant. The magnitudes of the coefficients in columns 3 and 4 are quite close, suggesting that the loss of statistical significance in column 4 is likely a result of reduced power - including fixed effects means that identification depends on within-individual variation, which is limited when many consumers have short search paths. These results are robust to alternative definitions of "surprisingly bad products", which are displayed in Appendix Tables A.1- A.3.

Dependent Variable:	$\|Price_{it} - Price_{it-1}\|$				$\|Pixel_{it} - Pixel_{it-1}\|$			
$BadProduct_{it-1}$	103.109***	100.824***	68.291***	62.003**	.418***	.411***	.362***	.165
	(26.614)	(26.560)	(24.102)	(24.978)	(.145)	(.145)	(.132)	(.135)
$SearchPercentile_{it}$	.	-93.074***	-62.885***	-54.060***	.	-.289***	-.179**	-.230***
	.	(17.209)	(15.630)	(16.402)	.	(.094)	(.086)	(.089)
$Constant$	170.110***	225.209***	980.316***	921.834***	1.756***	1.927***	5.132***	4.660***
	(4.964)	(11.328)	(27.681)	(34.994)	(.027)	(.062)	(.152)	(.190)
Density Controls	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes
N	6526	6526	6526	6526	6526	6526	6526	6526

Dependent Variable:	$\|Zoom_{it} - Zoom_{it-1}\|$				$\|Display_{it} - Display_{it-1}\|$			
$BadProduct_{it-1}$	.793**	.777**	.578*	.415	.101***	.100***	.102***	.083***
	(.368)	(.367)	(.326)	(.330)	(.019)	(.019)	(.017)	(.018)
$SearchPercentile_{it}$	.	-.645***	-.527**	-.335	.	-.029**	-.019*	-.024**
	.	(.238)	(.211)	(.217)	.	(.012)	(.011)	(.012)
$Constant$	3.113***	3.495***	10.519***	9.732***	.213***	.230***	.553***	.549***
	(.069)	(.157)	(.374)	(.463)	(.003)	(.008)	(.020)	(.025)
Density Controls	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes
N	6526	6526	6526	6526	6526	6526	6526	6526

Table 4: **Surprisingly Bad Products** Each panel records the results of regressions of absolute step size on an indicator for whether the last product was a "surprisingly bad product". Surprisingly bad products are defined as those which are searched by at least 10 individuals in the data but had a probability of purchase conditional on being searched of less than 5%. The additional controls are described in the text. Stars indicate statistical significance as folows: *** p<0.01 ** p<0.05 *p<0.1

The findings of these regressions are in line with what we would expect to observe if consumers made inferences about nearby products after each search. We interpret "surprisingly bad products" as those that yield, on average across individuals, lower utility than expected. The larger step size we observe in subsequent search is in line with consumers making the inference that nearby products are likely to yield similarly low utility, consistent with Proposition 2.

5 Structural Estimation

5.1 Econometric Specification

In order to take the model developed in Section 3 to the data on consumer search paths, we make several additional assumptions on the forms of the consumers' prior mean and covariance functions. We assume that consumers' prior means are linear in product characteristics:

$$\mu_i(X_j) = \alpha + X_j\beta_i \tag{5}$$

$$\beta_{ik} = exp(\eta_k + \nu_{ik}\sigma_{\beta k}) \tag{6}$$

$$\nu_{ik} \sim N(0,1) \tag{7}$$

where consumers have individual specific (random) coefficients β_i. Recall that X_j is a K dimensional vector of product attributes. We assume that each element β_{ik} of the coefficient vector $\beta_i = (\beta_{i1}, ..., \beta_{iK})$ is drawn independently across consumers i and product attributes k from a log-normal distribution with mean parameter $\eta_{\beta k}$ and standard deviation parameter $\sigma_{\beta k}$.

We assume that that consumers' prior covariance function $\kappa_i(X_j, X_l)$ is of the form given by equation (8). This is similar to the square exponential covariance function introduced earlier in the text but allows the covariance between $m_i(X_j)$ and $m_i(X_l)$ to decay with distance at different rates along different dimensions of the product characteristic space. In particular, there are K parameters ρ_k that control spatial correlation in utility along the K dimensions. The parameter λ

controls the overall variance level of the prior Gaussian process.

$$\kappa(X_j, X_l) = \lambda exp\left(\sum_{k=1}^{K} \frac{-\left(X_{jk} - X_{lk}\right)^2}{2\rho_k^2}\right) \tag{8}$$

To further simplify the consumer's problem, we impose that consumer i's cost of searching product j at period t, c_{ijt}, is given by equations 9 and 10, where μ_c and σ_c are parameters, and ζ_{ijt} is a logit error term that is drawn independently across t, i, and j. Consumers have individual specific cost parameters drawn from a log-normal distribution. The logit assumption simplifies subsequent computation.[6]

$$c_{ijt} = c_i + \zeta_{ijt} \tag{9}$$

$$\ln(c_i) \sim N(\mu_c, \sigma_c) \tag{10}$$

5.2 Parameter Normalization

The model is not identified without some normalization of the parameters. First, we normalize the scale of utility by setting $\sigma = 1$. We also normalize the level of utility by giving consumers an outside option with utility zero, setting $\hat{u}_{i0} = 0$ for all i. However, note that in our application to digital cameras we only observe an individual if they make at least one search. To deal with this, we assume that consumers must make at least one search (i.e there is no initial outside option), and afterwards can choose to stop searching without purchasing a product and obtain outside option utility $\hat{u}_{i0} = 0$.

Second, we normalize the relative value of additional search versus search costs by setting the overall variance parameter $\lambda = 1$. The intuition is that a scaling of λ raises the value of additional search, but so does a re-scaling of the search cost distribution, and it seems unlikely that these are separately non-parametrically identified.

Thus the parameters to be estimated comprise those determining the distribution of random co-

[6]Notice that the model nests a version of the standard random coefficients discrete model of Berry, Levinsohn, and Pakes (1995). When $\gamma = 0$ and $c_{ijt} = 0$ the model collapses to a probit choice model with linear utility and random coefficients.

efficients in the prior mean, $\{\{\eta_k\}_{k=1}^K, \{\sigma_{\beta k}\}_{k=1}^K, \alpha\}$, those determining the distribution of prior covariance functions, $\{\rho_k\}_{k=1}^K$, and the search cost parameters $\{\mu_c, \sigma_c\}$. For a K dimensional product characteristic space, there are therefore $3K+3$ parameters to be estimated. Let θ be the set of parameters to be estimated. Given this set of parameters, and a K dimensional vector of product attributes for each of the J products, the model generates a distribution of search paths.

5.3 Identification

We offer a heuristic argument for semi-parametric identification in three steps (non-parametric identification of the random coefficient distribution and search costs, keeping the Gaussian process structure parametric). Notice firstly that because consumers are modeled as using one-period look ahead, their first choice is simply the product with the highest prior mean utility. Consider then the problem of identifying the distribution of random coefficients from observing only the first choice made by consumers who face different choice sets (i.e. the set of products and their characteristics varies across consumers). Then this is a standard random coefficients model and dataset (a mixed multinomial probit), and from Berry and Haile (2014) we know that the model is identified.

In practice, observing repeated choices from the same individual will add very useful additional information, as in Berry, Levinsohn and Pakes (2004). Heterogeneity in preferences shows up in the data through correlation in product attributes across searches for the same individual. In particular, one can think of the variance parameters of the distribution of random coefficients, σ_β, as being identified by the amount of cross-individual variation in searched product attributes relative to within-individual variation. Table 5 breaks down the within and between individual variance for simulated search paths in a model with two product attributes. The first column records these variances, and the R^2 of a regression of searched attribute levels on individual fixed effects, for a model with $\sigma_\beta = 0$ for both products. In the second column, show the same objects when $\sigma_\beta = 4$. Increasing the variance of the random coefficients increases the between variance relative to the within variance, and increases the amount of variation in searched attributes across all individuals that can be explained by individual fixed effects. The intercept parameter, α, which controls the level of utility of all goods relative to the outside option is identified by the probability of a search path terminating without purchase.

The second part of the identification argument concerns the distribution of search costs. Different consumer types will have different expected payoff differences between their first and second choices. The probability of sampling their second choice (rather than stopping after a single choice) as a function of this expected payoff difference suffices to identify the distribution of search costs. Now, the expected payoff differences are latent, but the observed first choice provides of signal of the random coefficient and thus on the payoff differences. Formally, taking the distribution of random coefficients as known, for any choice set there is a known distribution of the benefit to additional search conditional on any first choice, as well as a probability of searching conditional on that choice. If the distribution of benefits induced by choice set and first choice variation were a set of point masses over the support of the search costs, this would immediately suffice for identification: one could invert the probability of search as each level of benefit c in the support of the search costs. In fact we face a mixture problem — we see the probability of search for different distributions of benefits — but it seems plausible that the search costs remain identified.

Finally, consider the parameters governing spatial correlation. Conditional on choosing to not to stop after sampling a single product, all options have the same search cost (up to logit error). Therefore the probability of sampling product k versus product l depends only on the difference in the prior part of utility and the difference in the option value coming from the Gaussian errors. Changing the covariance parameters will change the way in which consumers update their beliefs about this option value, and therefore change the incentives for consumers to explore the product space. Specifically, if consumers were ex-ante indifferent between all products, then when spatial correlation is high, sampling one product is very informative for close-by products, so consumers should either sample another close by product (if their experience was good) or sample a far away product (if their experience was bad). By contrast, when spatial correlation is low, the next product sampled should be uniformly distributed over the product space. So if the prior parts of utility were known, the spatial parameters would be identified from the rates at which search for k and l co-vary with the location of the initial product j. But since the priors are latent, again

Attribute		$\sigma_\beta = 0$	$\sigma_\beta = 4$
1	Within Individual Variance	1.038	1.023
	Between Individual Variance	0.216	0.388
	R^2 of regression on individual fixed effects	0.145	0.195
2	Within Individual Variance	3.805	3.730
	Between Individual Variance	1.021	1.698
	R^2 of regression on individual fixed effects	0.158	0.220

Table 5: **Identification of σ_β** Table records statistics on the distribution of product attributes within and across simulated search paths for two sets of parameter values. All parameters except $\sigma_{\beta 1} = \sigma_{\beta 2}$ are held constant. For each set of parameters, we simulate 10,000 search paths. Search is over a 2-dmensional attribute space with 50 products and product attributes randomly drawn from uniform $[0, 4]$ distributions. The other parameters are set as follows: $\mu_1 = -1$, $\mu_2 = 1$, $\alpha = -2$, $\lambda = 3$, $\rho_1 = 5$, $\rho_2 = 4$, $\mu_c = 0.8$, $\sigma_c = 0.2$. Within individual variance is the average across simulated paths of the variance within each path of a product attribute. Between individual variance is the variance across paths of the average attribute level within each path.

one would need to take into account selection into the initial product j and make use of variation in that product and the choice sets more generally.

5.4 Estimation

We estimate the model by constructing a likelihood function on the observed consumer search paths and choices. Under the assumption that search costs are given by equation (9) with logit errors, the probability of a consumer choosing to search product $j \in \tilde{\boldsymbol{J}}$ conditional on being at state $\mathcal{S}$, but unconditional on the realizations of the logit cost shocks is given by:

$$\tilde{P}_i(j|\mathcal{S}) = \frac{exp\left(E[\max\{\hat{u}, u_j\}|\mathcal{S}] - c\right)}{exp(\hat{u}) + \sum_{l \in \tilde{\boldsymbol{J}}} exp\left(E[\max\{\hat{u}, u_l\}|\mathcal{S}] - c\right)} \tag{11}$$

Suppose consumer i searches T_i times before stopping. Let j_{it} be the tth product searched. Let $j_{it} = 0$ indicate stopping and purchasing the highest utility sampled product (or the outside option). Finally, let $\hat{j}_i$ indicate the product purchased. If the consumer's state variable, $\mathcal{S}$, was fully observable to the econometrician, the likelihood of the consumer's search path would then be given by equation 12.

$$L_i(\{j_{it}\}_{t=0}^{T_i}, \hat{j}_i | \{\mathcal{S}_t\}_{t=0}^{T_i}, \theta, \nu_i, c_i) = \left(\prod_{t=0}^{T_i - 1} \tilde{P}_i(j_{it} | \mathcal{S}_t) \right) \tilde{P}_i(0 | \mathcal{S}_{T_i}) 1 \left(u_{\hat{j}_i} = \hat{u}_{j_{iT_i}} \right) \tag{12}$$

Since the econometrician does not observe the utility draws that enter $\mathcal{S}$, it is necessary to integrate them out of the likelihood function. Conditional on the consumer's utility function, $m_i(X_j)$, the utility u_{ij} is distributed according to $u_{ij} \sim N(m_i(X_j), 1)$. There are T_i integrals inside the inner parentheses of equation 13 corresponding to each utility draw observed by consumer i. Each integral is an expectation with respect to distributions $F_{ji}(u) = \Phi\left(u_j - m_i(X_j)\right)$, where $\Phi(\cdot)$ is the standard normal CDF.

Since the econometrician does not observe $m_i(X_j)$ either, it is necessary to take the expectation of the likelihood over draws of $m_i(X_j)$ from consumer i's prior Gaussian process. This is represented by the outer integral in equation 13.

$$L_i(\{j_{it}\}_{t=0}^{T_i}, \hat{j}_i | \theta, \nu_i, c_i) = \int \left(\int ... \int L_i(\{j_{it}\}_{t=0}^{T_i}, \hat{j}_i | \{\mathcal{S}_t\}_{t=0}^{T_i}, \theta, \nu_i, c_i) dF_{ji}(u_{j_{i0}}) ... dF_{ji}(u_{j_{iT_i - 1}}) \right) dGP(\mu_i, \kappa_i) \tag{13}$$

Finally, note that the likelihood function in equation 13 depends on the individual specific terms c_i and ν_i - the random coefficients that determine consumer i's prior Gaussian process. To write the likelihood in terms of only the parameters to be estimated, θ, these random coefficients must be integrated out. Equation 14 is the resulting likelihood of individual i's search path and purchase, conditional only on the parameters to be estimated, θ.

$$L_i(\{j_{it}\}_{t=0}^{T_i}, \hat{j}_i | \theta) = \int \int L_i(\{j_{it}\}_{t=0}^{T_i}, \hat{j}_i | \theta, \nu_i, \omega_i)) dF_\nu(\nu_i) dF_c(c_i) \tag{14}$$

Our estimation procedure uses this likelihood to form a Monte Carlo Markov Chain from arbitrary starting values using a version of the Metropolis-Hastings algorithm with flat priors. Draws from this Markov Chain converge to the posterior distribution implied by the data. For each parameter, we take the mean of this distribution to be our parameter estimate.

	Estimate	SE		Estimate	SE
η_1 (log price)	-0.683	0.032	ρ_1 (log price)	2.151	0.116
η_2 (display)	0.292	0.010	ρ_2 (display)	1.710	0.354
α	5.060	0.270			
$\sigma_{\beta 1}$ (log price)	0.599	0.030	c	1.627	0.012
$\sigma_{\beta 2}$ (display)	0.133	0.006	σ_c	0.250	0.012

Table 6: **Estimated Parameters** Table records the results of estimating the model on the digital camera data, comprising 966 search paths. We include only log price and pixel size as ex ante observable attributes in the estimated model. Estimation uses 5,000 MCMC draws, with the first 1,000 used for burn-in and dropped from reported estimates. The parameter estimates are the mean of the MCMC posterior distribution, and the standard deviations recorded are the standard deviations of the posterior distribution.

6 Results

6.1 Model Fit

We estimate our model on the digital camera search path data from BKM using the MCMC approach discussed above. Because of computational limitations, we restrict the model to include only two product attributes - log price and display size. When testing models with additional product attributes we found that the estimated utility parameters on price and display were significantly larger than the other utility parameters.

The estimated parameters are presented in Table 6. As we might expect, the mean of the random coefficient on log price is negative and significant, and the mean of the random coefficient on display size is positive and significant. The variance of the Gaussian process from which consumers' preferences are drawn, λ, and the covariance parameters ρ_k for the two attribute dimensions are positive and significant. Recall that as $\rho_k \to 0$, the model converges to a standard sequential search model without learning. The statistically significant estimates of ρ_k suggest that a model in which the covariance of the Gaussian process is 0 - a standard sequential search model without spatial learning in the style of Weitzman (1979) - would be rejected in favor of the current model. Since we can reject the hypothesis that $\rho_k = 0$ in favor of the alternative $\rho_k > 0$, the data on search paths provides evidence that spatial learning is taking place, and that consumers update their beliefs

	Data		Simulations	
	Mean	SD	Mean	SD
Search Length	5.299	6.466	5.436	7.187
Chosen Product Discovered	0.790	0.265	0.718	0.295
Within Search Log Price Variance	0.480	0.306	0.886	0.418
Within Search Display Variance	0.236	0.148	0.382	0.224
Distance of First Search to Purchase in Log Price	0.359	0.545	0.633	0.838
Distance of First Search to Purchase in Display	0.165	0.258	0.288	0.434

Table 7: **Model Fit** Table records the mean and standard deviation of various statistics across search paths in the data, and across 10,000 search paths simulated using the products in the data and the estimated parameters. Search path length is the number of unique products viewed. The second row records the search percentile, as defined in Section 3, at which the product eventually chosen is first viewed

about un-searched objects as they search. Note that although it is tempting to interpret the ρ_k parameters as indicating that utility draws are more correlated along the price dimension than the display dimension, the relative magnitudes of the ρ_k parameters cannot be directly interpreted. In particular, the parameters depend on, for example, the units of measurement of each attribute.

Note that the estimated coefficient on price cannot be used to give a dollar interpretation to the estimated search costs since the coefficient on price includes both the direct effect of price on utility and the indirect effect of price on consumers' prior beliefs about quality. Furthermore, our estimation approach which imposes a one period look ahead strategy rather than solving for the optimal forward looking policy will also bias estimated the search cost. Since the one period continuation value is a lower bound on the true continuation value, the estimated search cost should be *lower* than the true cost in order to rationalize observed search.

Table 7 illustrates the fit of the model to the data. The first two columns record the mean and standard deviation of various search path moments across search paths in the data. The third and fourth column record these same moments across 10,000 search paths simulated using the estimated parameters. In particular, for each simulated path i, a set of random coefficients and costs, $\{\beta_{1i}, \beta_{2i}, c_i\}$ are drawn from the distributions implied by the estimated parameters. The parameters $\{\beta_{1i}, \beta_{2i}\}$ along with the covariance parameters define the Gaussian process from which the consumer's ex-ante unknown preferences are drawn, which also defines the consumer's ex-ante beliefs. We take one draw from this Gaussian process and then simulate the search path of a consumer with these preferences (this draw is the $m_i(X)$ in equation 1).

The results in the first two rows indicate that the distribution of search path lengths, and the search percentile at which the purchased product is first discovered in the simulated paths match the data closely. As discussed in Section 5, for any values of the other parameters, the distribution of search path lengths can be matched by setting the distribution of search costs. Since the mapping from search cost to search length is clear, it should not be surprising that this moment is matched well by the model, despite the several simplifications we make in estimation. The close match between the data and simulations of the percentile at which the chosen product is first discovered is harder to attribute to a particular parameter. The remaining rows record within search attribute variance and the distance in attribute space from the first search to the purchased product. These statistics indicate that the estimated parameters induce significantly more variance in searched product attributes than there is in the data. One possible reason for this result is that the distribution of random coefficients is not sufficiently flexible to capture the true distribution of preferences across consumers. As discussed by Dubé et al. (2010), assuming independent, normally distributed random coefficients can lead to a finding of "excess inertia" in repeated consumer choice.

6.2 Manipulating Search Rankings Under Spatial Learning

One of the innovative features of our model of search is the introduction of path dependence. What a consumer learns from search determines what she searches next, and ultimately, what she purchases. If consumers could be manipulated to begin their search paths at different products, their search paths and purchases would be different. Consider an experiment in which all consumers are forced to view a particular product before beginning their search through the remaining products. Changing this start position will change the beliefs consumers have at the beginning of their search, and therefore change their subsequent paths. This stands in contrast to models of sequential search without correlated learning, in which manipulating the first object searched has no effect on the sequence of objects searched thereafter, only on the point in the sequence at which the consumer stops searching. This also implies that in a model without learning, a consumer who purchased product A would purchase either A or B in a counterfactual world where he is forced to view B first, whereas in a model with learning, forcing a consumer to view product B first could alter their search path such that they end up purchasing some third product C. For instance, if a consumer learns that they would obtain an unexpectedly high payoff from B, they might search through other products that are similar to B and yield similarly high payoffs, and end up purchasing such a product, C, that they would not have searched at all had they been free to start their search anywhere.

This "forced search" experiment can be thought of as modeling the effects of search ranking or sponsored search results in online retail. Search engines and online retail platforms such as Amazon, Google, and eBay frequently place sponsored products or advertisements at the top of search results pages. Although we do not model the effect of search ranking on the decision to search a particular object, it is well documented that placing items near the top of results page increases the frequency with which those items are viewed. We simulate a counterfactual in which the platform can force consumers to view a particular item first, rather than simply encourage them through search rankings.[7] We add a new product which all consumers are forced to view first before continuing to search. The first search after viewing this new product is costless to the consumer. The experiment therefore consists of providing consumers with information about a particular product before allowing them to search as normal.

To illustrate how search paths and consumer utility can be affected by this type of search manipulation, we take an example in which all consumers have a low value of ϵ_{ij} for the new product j. We assume that new product j has ϵ_{ij} composed of a part that is common to all consumers and an idiosyncratic part. In particular, $\epsilon_{ij} = \xi_j + \delta_{ij}$ where $\xi_j = -5$ and $\delta_{ij} \sim N(0, \sqrt{0.5})$. We imagine that this particular value of γ_j is a draw from a mean zero normal distribution with variance $\sqrt{0.5}$, so that from the consumer's perspective, $\epsilon_{ij} \sim N(0, 1)$, as in the estimated model.

We simulate 10,000 search paths under this counterfactual, and compare the distribution of search

[7]Note that in this counterfactual, we imagine that all products are viewed on one platform or website, rather than across multiple domains.

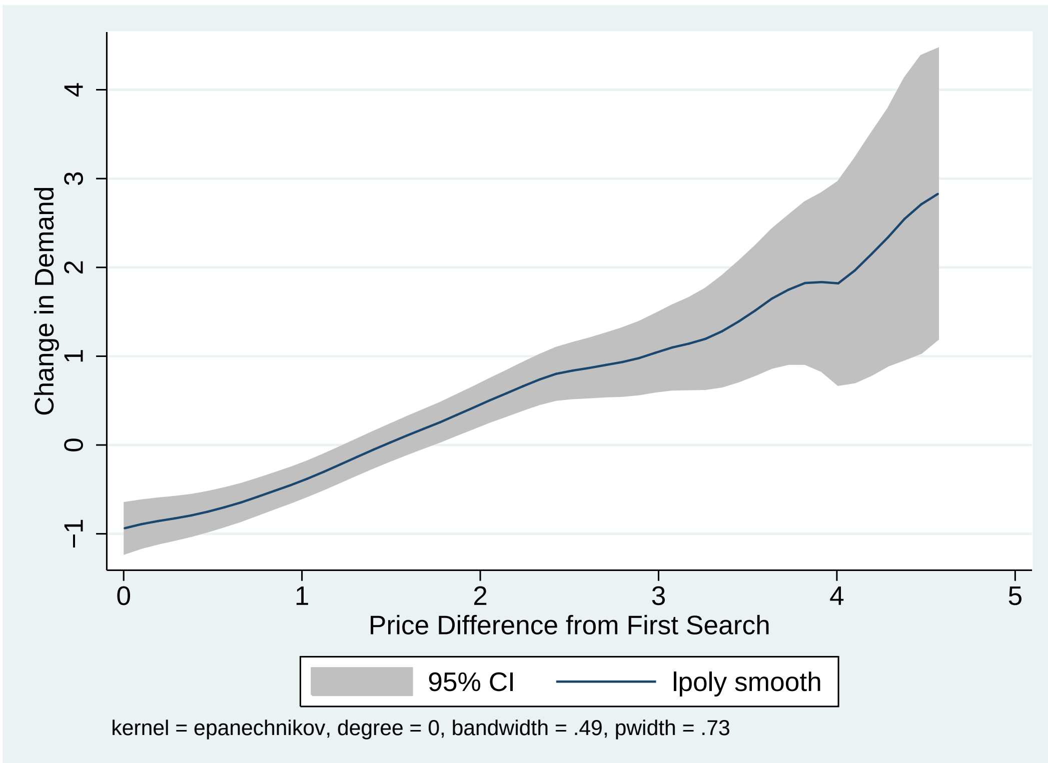


Figure 9: **Substitution from Manipulation of Search Rankings** For each product in the digital camera data, the x-axis records the difference in price between that product and the counterfactual product that consumers are made to view first before beginning their search. The counterfactual product has log price of 4 and display size of 2.5. The y-axis records the change in simulated demand per 1000 search paths between the counterfactual "forced search" simulation and the baseline simulation. Negaive values mean that that demand is lower under the counterfactual simulation. The solid line is a kernel regression using an Epanechnikov kernel, and the shaded area is 95% confidence interval.

paths to the distribution of paths when the model is simulated without search manipulation. Because consumers are made to view a product with a particularly low value of ϵ_{ij}, they will infer, probably incorrectly, that nearby products will also yield low utility. We have intentionally constructed the counterfactual so that consumers are forced to view a *misleading* product before continuing to search. Figure 9 shows the change in demand between the baseline and counterfactual simulations. The y-axis records the change in the number of purchases per 1000 search paths between the two simulations, and the x-axis records the absolute difference in log price between each product and the "new product" that consumers are forced to view first. Demand falls for products that are close to the new product in price and rises for products that are further away. We would not observe this pattern of substitution in a model without spatial learning. Recall that in such a model, any substitution in this counterfactual would be towards the new product that consumers are forced to view first - the demand for all other products would remain the same or fall. The information provided by the first product can only be *misleading*, and alter the consumer's subsequent search path, insofar as consumers' beliefs are correlated across products.

Table 8 records some additional statistics on search paths in the two simulations. The share of consumers who terminate their search by choosing the outside option increases from 12% to 17% when the forced search counterfactual is imposed. It is clear that the additional information provided in the counterfactual *reduced* the consumption utility of those additional consumers who chose not to purchase a camera. In fact, the difference in mean consumption utility (that is, excluding the contribution of search costs to utility) between the counterfactual and baseline simulations, recorded in row 4 of Table 8, is negative. The model therefore captures the idea that providing additional, free information about a particular products can reduce consumer welfare if that information is misleading about the payoffs from other products. In a model without spatial learning, providing consumers with additional information does not have this effect since there is no path dependency in search. Row 5 records the average consumption utility for analogous simulations with spatial learning tuned off. that is, where the covariance between $m_i(X)$ and $m_i(X')$ is 0 for and two products X and X'. In this case, the average utility in the "forced search" counterfactual is almost identical to the average utility in the baseline simulation.

	Free Search	Forced Search Counterfactual
Search Length	5.299	5.910
Outside Option Share	12%	17%
Mean Consumption Utility: Learning Model	3.956	3.751
Mean Consumption Utility: No Learning Model	4.232	4.235

Table 8: **Manipulation of Search Rankings** Table records statistics on simulated search paths. The first column uses the estimated parameters and the products in the digital camera data. The second column uses the same parameters and product, but forces consumers to search a new product with log price of 4 and display size of 2.5 before commencing search, as described in the text. Both columns use 10,000 simulated paths. Outside option share is the number of consumers who choose not to purchase any product. Mean consumption utility is utility excluding the cost of search. The final row uses simulations from a model without spatial learning, as described in the text.

The path dependency introduced by spatial learning therefore complicates the effects of manipulating search rankings on demand. A useful direction for future research would be to provide direct empirical support for these effects using data on search paths with exogenous changes in search rankings. Spatial learning also has related supply-side implications for optimal product positioning and investment in quality that we would like to investigate in future.

7 Conclusion

In this paper, we develop a model of search with spatial learning and investigate the implications on platform power in online retail. Consumers are initially uncertain of the utility yielded by the set of available products, which they learn abou through search. Searching a particular product not only provides information about that product, but provides a signal about how much the consumer is likely to value similar products - those that are "nearby" in product attribute space.

In a simulated example we demonstrate how spatial learning induces path dependence in search. In standard models of sequential search, the consumer's choice of *whether* to continue searching depends on the observed utilities of those objects already searched. With spatial learning, the decision of *which* product to search also depends on past observations. For example, a negative observed utility for a particular product is likely to direct subsequenct search away from similar products. We establish some simple coparative statica on the consumer's "search ranking" of products under a one period look ahead assumption that formalize this intuition.

We then document stylized facts using data on consumer search paths in online search for digital cameras that provide support for the intuitive notion that consumers are exploring the product space to learn about their preferences, and that learning about the payoff from one object provides information on the payoffs from similar objects. Consumers initially take large steps over a wide range of the product attribute space before focusing on products close to the ultimately purchased product in later search. Consumers also take larger steps in attribute space away from the last product searched when that product is "surprisingly bad," in the sense that it is frequently searched but rarely purchased.

We then make additional assumptions that allow the parameters of the search model to be estimaed using the search path data. We use the estimated model to investigate the implications of this type of spatial learning on the effects of manipulating search rankings in online retail. The path dependency induced by this type of learning means that forcing consumers to start their search with a particular product changes the order of products subsequently searched. We show that when the product that consumers are made to view first yields an unexpectedly, and misleadingly, low utility, consumers tend to substitute away from products which are near this first product in attribute space. We show that this effect means that, unlike in a model without spatial learning, providing additional information to consumers can reduce consumer welfare on average. These results suggest other interesting effects that might be studied using this model. The presence of spatial learning alters firms' incentives to invest in quality, and may alter optimal product positioning for firms entering a market. For example, spatial learning will reduce the expected profit from introducing a product close in attribute space to another product which is perceived as low quality. Further investigation of these supply side implications is a promising avenue for future research.

References

Adam, Klaus, "Learning While Searching for the Best Alternative," *Journal of Economic Theory*, 2001, *101* (1), 252–280.

Berry, Steven and Phillip Haile, "Identification in Differentiated Product Markets Using Market Level Data," *Econometrica*, 2014, *82* (5), 1749–1797.

____, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–90.

____, ____, and ____, "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 2004, *112* (1), 68–105.

Boik, Andre, Shane Greenstein, and Jeffrey Prince, "The Empirical Economics of Online Attention," Working Paper 22427, National Bureau of Economic Research July 2016.

Bronnenberg, Bart J., Jun B. Kim, and Carl F. Mela, "Zooming In on Choice: How Do Consumers Search for Cameras Online?," *Marketing Science*, 2016, *35* (5), 693–712.

Crawford, Gregory and Matthew Shum, "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 2005, *73* (4), 1137–1173.

Dickstein, Michael, "Efficient Provision of Experience Goods: Evidence from Antidepressant Choice," Working Paper, Harvard University 2014.

Dinerstein, Michael, Liran Einav, Jonathan Levin, and Neel Sundaresan, "Consumer Price Search and Platform Design in Internet Commerce," Working Paper 20415, National Bureau of Economic Research August 2014.

Honka, Elisabeth and Pradeep Chintagunta, "Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry," *Marketing Science*, 2017, *36* (1), 21–42.

Kim, Jun B., Paulo Albuquerque, and Bart J. Bronnenberg, "Online Demand Under Limited Consumer Search," *Marketing Science*, 2010, *29* (6), 1001–1023.

Koulayev, Sergei, "Search for differentiated products: identification and estimation," *RAND Journal of Economics*, September 2014, *45* (3), 553–575.

McCall, J. J., "Economics of Information and Job Search," *The Quarterly Journal of Economics*, 1970, *84* (1), 113–126.

Osborne, Michael A., Roman Garnett, and Stephen J. Roberts, "Gaussian processes for global optimization," in "LION" 2009.

Rasmussen, Carl Edward and Christopher K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*, The MIT Press, 2005.

Rothschild, Michael, "Searching for the Lowest Price When the Distribution of Prices Is Unknown," *Journal of Political Economy*, 1974, *82* (4), 689–711.

Santos, Babur De Los, Ali Hortaçsu, and Matthijs R. Wildenbeest, "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, May 2012, *102* (6), 2955–80.

Simon, Herbert A., *Designing Organizations for an Information-Rich World*

Ursu, Raluca, "The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions," Working Paper, NYU Stern 2017.

Varian, Hal R., "A Model of Sales," *The American Economic Review*, 1980, *70* (4), 651–659.

Weitzman, Martin, "Optimal Search for the Best Alternative," *Econometrica*, 1979, *47* (3), 641–54.

Appendix

Table A.1: Surprisingly Bad Products: Alternative Definition 1

Dependent Variable:	$\lvert Price_{it} - Price_{it-1} \rvert$				$\lvert Pixel_{it} - Pixel_{it-1} \rvert$			
$BadProduct_{it-1}$	46.848***	43.730**	18.858	15.474	.300***	.291***	.260***	.094
	(18.147)	(18.118)	(16.442)	(17.120)	(.099)	(.099)	(.090)	(.093)
$SearchPercentile_{it}$	.	-92.785***	-62.968***	-54.289***	.	-.285***	-.175**	-.229**
	.	(17.227)	(15.643)	(16.419)	.	(.094)	(.086)	(.089)
$Constant$	170.021***	225.114***	982.663***	923.624***	1.747***	1.916***	5.122***	4.659***
	(5.083)	(11.417)	(27.733)	(35.040)	(.028)	(.062)	(.152)	(.190)
Density Controls	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes
N	6526	6526	6526	6526	6526	6526	6526	6526

Dependent Variable:	$\lvert Zoom_{it} - Zoom_{it-1} \rvert$				$\lvert Display_{it} - Display_{it-1} \rvert$			
$BadProduct_{it-1}$	.121	.099	.078	-.117	.064***	.063***	.064***	.049***
	(.251)	(.251)	(.222)	(.226)	(.013)	(.013)	(.012)	(.012)
$SearchPercentile_{it}$	.	-.650***	-.530**	-.345	.	-.028**	-.019*	-.023**
	.	(.238)	(.211)	(.217)	.	(.012)	(.011)	(.012)
$Constant$	3.131***	3.517***	10.550***	9.771***	.212***	.228***	.552***	.548***
	(.070)	(.158)	(.375)	(.463)	(.004)	(.008)	(.020)	(.025)
Density Controls	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes
N	6526	6526	6526	6526	6526	6526	6526	6526

Notes: Each panel records the results of regressions of absolute step size on an indicator for whether the last product was a “surprisingly bad product”. Surprisingly bad products are defined as those which are searched by at least 10 individuals in the data but had a probability of purchase conditional on being searched of less than 10%. The additional controls are described in the text. Stars indicate statistical significance as folows: *** p<0.01 ** p<0.05 *p<0.1

Table A.2: Surprisingly Bad Product: Alternative Definition 2

Dependent Variable:	$\|Price_{it} - Price_{it-1}\|$				$\|Pixel_{it} - Pixel_{it-1}\|$			
$BadProduct_{it-1}$	50.387***	50.348***	39.722***	44.642***	.129*	.129*	.116	.043
	(14.420)	(14.388)	(13.051)	(13.684)	(.079)	(.079)	(.072)	(.074)
$SearchPercentile_{it}$	.	-94.084***	-63.540***	-54.621***	.	-.293***	-.182**	-.232***
	.	(17.210)	(15.627)	(16.392)	.	(.094)	(.086)	(.089)
$Constant$	167.057***	222.678***	978.896***	917.341***	1.753***	1.927***	5.140***	4.662***
	(5.235)	(11.436)	(27.703)	(35.039)	(.029)	(.063)	(.152)	(.190)
Density Controls	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes
N	6526	6526	6526	6526	6526	6526	6526	6526

Dependent Variable:	$\|Zoom_{it} - Zoom_{it-1}\|$				$\|Display_{it} - Display_{it-1}\|$			
$BadProduct_{it-1}$	.286	.286	.381**	.292	.033***	.033***	.030***	.018*
	(.199)	(.199)	(.176)	(.181)	(.010)	(.010)	(.009)	(.010)
$SearchPercentile_{it}$	.	-.653***	-.533**	-.339	.	-.030**	-.020*	-.025**
	.	(.238)	(.211)	(.217)	.	(.012)	(.011)	(.012)
$Constant$	3.103***	3.489***	10.500***	9.704***	.212***	.230***	.556***	.550***
	(.072)	(.158)	(.374)	(.464)	(.004)	(.008)	(.020)	(.025)
Density Controls	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes
N	6526	6526	6526	6526	6526	6526	6526	6526

Notes: Each panel records the results of regressions of absolute step size on an indicator for whether the last product was a "surprisingly bad product". Surprisingly bad products are defined as those which are searched by at least 5 individuals in the data but had a probability of purchase conditional on being searched of less than 5%. The additional controls are described in the text. Stars indicate statistical significance as folows: *** $p<0.01$ ** $p<0.05$ *$p<0.1$

Table A.3: Surprisingly Bad Products: Alternative Definition 3

Dependent Variable:	$\|Price_{it} - Price_{it-1}\|$				$\|Pixel_{it} - Pixel_{it-1}\|$			
$BadProduct_{it-1}$	39.334***	38.268***	25.008**	28.780**	.155**	.152**	.137**	.043
	(12.826)	(12.800)	(11.616)	(12.254)	(.070)	(.070)	(.064)	(.066)
$SearchPercentile_{it}$	.	-93.323***	-63.042***	-53.990***	.	-.290***	-.180**	-.231***
	.	(17.216)	(15.634)	(16.403)	.	(.094)	(.086)	(.089)
$Constant$	166.795***	222.149***	979.618***	918.474***	1.743***	1.916***	5.127***	4.660***
	(5.373)	(11.533)	(27.758)	(35.092)	(.029)	(.063)	(.152)	(.190)
Density Controls	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes
N	6526	6526	6526	6526	6526	6526	6526	6526

Dependent Variable:	$\|Zoom_{it} - Zoom_{it-1}\|$				$\|Display_{it} - Display_{it-1}\|$			
$BadProduct_{it-1}$	.103	.095	.206	.074	.035***	.034***	.032***	.019**
	(.177)	(.177)	(.157)	(.162)	(.009)	(.009)	(.008)	(.009)
$SearchPercentile_{it}$	.	-.651***	-.529**	-.338	.	-.029**	-.020*	-.024**
	.	(.238)	(.211)	(.217)	.	(.012)	(.011)	(.012)
$Constant$	3.122***	3.509***	10.514***	9.739***	.211***	.228***	.553***	.549***
	(.074)	(.160)	(.375)	(.464)	(.004)	(.008)	(.020)	(.025)
Density Controls	No	No	Yes	Yes	No	No	Yes	Yes
Individual FE	No	No	No	Yes	No	No	No	Yes
N	6526	6526	6526	6526	6526	6526	6526	6526

Notes: Each panel records the results of regressions of absolute step size on an indicator for whether the last product was a "surprisingly bad product". Surprisingly bad products are defined as those which are searched by at least 5 individuals in the data but had a probability of purchase conditional on being searched of less than 10%. The additional controls are described in the text. Stars indicate statistical significance as folows: *** p<0.01 ** p<0.05 *p<0.1