Consumer Search with Spatial Learning*

Charles Hodgson and Greg Lewis†

March 27, 2017

Abstract

We introduce a model of spatial learning and search by imperfectly informed consumers with unit demand. Sampling one product causes them to update not only their payoff to choosing that product but also their beliefs about the payoff of 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. Using data on online camera purchases, we argue that consumer search is indeed spatial, as consumers who sample unexpectedly low quality products tend to subsequently sample products that are far away in attribute space. We develop a flexible parametric specification of the model where consumer utility is sampled as a Gaussian process and use it to estimate demand in the camera data using Markov Chain Monte Carlo (MCMC) methods. We conclude with a counterfactual experiment in which we manipulate the initial product shown to a consumer, finding that a bad initial experience can lead to early termination of search. Product search rankings can therefore substantially affect the consumer search path and their eventual purchase decision.

1 Introduction

Consumer search is typically modeled as a setting in which a consumer must pay a cost to learn the payoffs associated with various objects before choosing which to consume. This paper starts with the observation that, in many real life settings, learning about the payoff from one object should change the consumer’s beliefs about the payoff from other, similar objects. For instance, suppose you are shopping for a new car, and think you might like to buy a particular small hatchback. You take the car for a test drive and decide that it is too small for you. It would be surprising if you then went on to test drive another small hatchback. What you learned from test driving the car was not only that purchasing this particular car would yield a low payoff, but that you don’t like this car because of

---

*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, and Tim Bresnahan for helpful comments.

†Hodgson: Department of Economics, Stanford University. Lewis: Microsoft Research and NBER.
its size. You would sensibly infer that the payoff from other, similarly sized cars is likely to be low. What you learn about the car you tested is generalizable to other vehicles because you believe that the mapping from attributes to preferences is correlated in attribute space: cars of a similar size are likely to yield a similarly low payoff.

In this paper we build a model of consumer search that captures this intuition. Consumers face a set of multi-attribute products, with some attributes known to the consumer in advance of search. Consumers have uncertain prior beliefs about the mapping from these ex-ante known product attributes to utility, and learn about this mapping through search. To capture the notion that information gathered through the search is informative about un-searched objects, we assume consumers’ utility functions are drawn from a Gaussian process on the product attribute space. For any value of the ex ante known attributes, consumers therefore have normally distributed beliefs over their true utility. For any two products, consumers’ beliefs over utility are a bivariate normal distribution with covariance determined by the distance between the two objects in attribute space. After observing their utility for a particular object by searching it, consumers update their beliefs over the utility of all other objects. Beliefs are updated more for un-searched objects that are close to the searched object in attribute space.

This model can be interpreted in two ways. In one interpretation, consumers do not know exactly how much they value different combinations of attributes, but believe, correctly, that their true payoff function is a continuous function in these attributes drawn from a Gaussian process with a linear mean. In the other, consumers have linear utility over the ex ante known attributes, but do not know the utility value of other ex-ante unobserved attributes which is drawn from a mean zero Gaussian process defined over the observed attributes. For example, a consumer searching for a digital camera online might know he prefers a camera with a larger zoom. Clicking through to a product page, he might read customer reviews and learn that cameras with particularly large zooms are heavy and difficult to carry. The consumer would then update his beliefs by reducing his expected utility for high zoom cameras. One could interpret this as a change in how the consumer values zoom, or as a change in the consumer’s beliefs about other products attributes (e.g. weight) that are correlated with zoom.

This modeling approach is closely linked to the literature on Gaussian processes in machine learning, as summarized in Rasmussen and Williams (2006). In this literature, optimal Bayesian learning based on Gaussian processes is used to construct algorithms for maximizing unknown functions subject to some constraints. The algorithms designed in this literature are attempts to create practical solutions to learning problems very similar to those faced by a consumer searching for the optimal product. The observation that assuming a Gaussian process is a tractable way to design machine learning algorithms applies equally to the modeling of learning by rational agents. Studies in economics including Covert (2015) have adapted Gaussian process learning from this literature as a nonparametric estimation technique. However, to our knowledge this is the first paper in economics to build a model of consumer learning in which economic agents are endowed with Gaussian process priors about their utility.

To motivate our modeling choices, we present a series of stylized facts describing patterns in search paths using data
on online search for digital cameras collected by Bronnenberg, Kim, and Mela (2016). 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 attributes searched declines as search progresses. To these we add the new findings that the difference in attributes between successive searches declines over the search path, and that consumers tend to take larger steps in attribute space after viewing rarely purchased products. We argue that existing models of sequential search cannot rationalize these patterns, and suggest that a model with spatial learning is one way to account for them. We use the same data to estimate the parameters of the model, noting that the learning behavior captured by the model is identified by the path of search.

The path dependence induced by spatial learning has implications for the design of online retail platforms. We use a counterfactual simulation to demonstrate how forcing consumers to view a particular product first affects the path of search and the resulting demand. We take this experiment to represent the effect of manipulating search rankings, for example through the placement of sponsored results at the top of the page, on consumers' search behavior. The results of this experiment demonstrate that, under spatial learning, providing consumers with additional information on a particular product can reduce welfare insofar as it results in consumers' making incorrect inferences about other products.

1.1 Related Literature

Consumer search is typically cast as a problem of price discovery. In the classic model of sequential search developed by McCall (1970), consumers with unit demand draw alternatives from a distribution of prices or utilities until the payoff from consuming the best alternative drawn so far outweighs the expected payoff from searching once more. This model has proved influential, and particularly useful in explaining various economic phenomena, such as price dispersion (see for example, Varian, 1980), as the equilibrium outcome of a market in which consumers have limited information. However, this model has nothing to say about the order in which products are sampled, since all products are ex ante identical to the consumer. As detailed data on consumer search paths, for example the browsing data described by Bronnenberg, Kim, and Mela (2016), becomes more widely available, more realistic models of the search process are required to rationalize observed behavior. We develop such a model by building on two strands of the consumer search literature that followed McCall (1970).

First, in order for a model to say anything about the path of search, the objects being searched through cannot be ex-ante identical to the consumer. One of the first papers to allow for ex ante differentiable search alternatives was Weitzman (1979). In his model, Weitzman allows the mean and standard deviation of the distribution of utility yielded by each search alternative, or “box”, to differ. He shows that the consumer’s optimal strategy is to search through boxes in the order of the reservation utilities implied by the distribution associated with each box. As we note below, our model is a strict generalization of a version of Weitzman’s model.
The second strand of the search literature that is relevant to our model is the work following Rothschild (1974) on search with learning. Typically, these models involve consumers updating their beliefs about the distribution from which searched objects are drawn, which means that consumers update their reservation value as they search. Recent papers in this literature include De ls Santos, Hortacsu, and Wildenbeest (2016) and Koulayev (2013). A handful of papers combine consumer learning with ex ante differentiable search alternatives. For example, Adam (2001) derives the consumer’s optimal policy in a setting with multiple types of search alternatives, each associated with a different ex ante unknown utility distribution. These papers are also related to the literature on experience goods, for example Crawford and Shum (2005) and Dickstein (2014). The experience goods problem has a similar structure to the consumer search problem, with the difference that in experience goods models, consumers receive a consumption payoff every time they sample an object. Dickstein’s (2014) model is particularly close to ours. In his model, beliefs about utility are correlated across alternatives within nests. Our contribution to this literature on top of the models of Adam (2001) and Dickstein (2014) is to allow beliefs to be correlated across continuous characteristics - not simply within discrete nests.

The remainder of the paper proceeds as follows. Section 2 describes the data on consumer search paths we use to test our model, and presents stylized facts from this data that motivate a model of spatial learning. Section 3 outlines the model. Section 4 describes the estimation and identification of the model using data on search paths. Section 5 presents the results of the estimation, and section 6 concludes.

2 Consumer Search: Stylized Facts

2.1 Data

To motivate our model of consumer search as a process of spatial learning we report stylized facts describing patterns in consumer search using online browsing data. We use 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 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.\(^1\) 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 1 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.\(^2\) Table 2 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 \(n\)th product is \(\frac{n}{N}\). Note that the \(N\)th product is not necessarily the product purchased. The chosen product is typically discovered towards the end of search. The lower panel of Table 2 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 is defined analogously. Notice that the standard deviation of price for products searched is larger than for products purchased, and the 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.

---

1. In later revisions we hope to make use of variation in price over time and across retailers.
2. 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.
2.2 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 1 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 that 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 progresses. Figure 2, 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 n\textsuperscript{th} 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 1: 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 3 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 n\textsuperscript{th} search has a step size in price of \( \frac{|\text{price}_{n-1} - \text{price}_n|}{\text{meanpricediff}} \) where \( \text{price}_n \) is the price of the consumer’s n\textsuperscript{th} searched product, and \( \text{meanpricediff} \) is the average absolute difference in prices across all pairs of products. The x-axis records search percentile, as in Figure 1. 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 1, 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 1, 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 2 and 3, 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 we develop in Section 3 builds on the idea that when consumers view one product they make inferences about their preferences for similar products, and that 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. This hypothesis is 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 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 3 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.

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. These results, as well as those documented by Figures 1-3, motivate the model of search with spatial learning about preferences that we present in Section 3 below. In this model, the path that consumers take through attribute space as they search is determined by their prior beliefs over the mapping from attributes to utility and the observations they make as they search which allow them to update their beliefs about this mapping.
3 Model

3.1 Consumer’s Problem

Consumers, indexed by \( i \), with unit demand search sequentially through a finite set \( J \) of available products. Products, indexed by \( j \), are described by a vector of characteristics \( X_j \in X \subseteq \mathbb{R}^K \). Consumer \( i \) obtains a utility \( u_{ij} \in \mathbb{R} \) from consuming product \( j \) given by:

\[
u_{ij} = m_i(X_j) + \epsilon_{ij}\]

(1)

Where \( m_i : X \rightarrow \mathbb{R} \) is a function that maps a vector of characteristics to mean utility, and \( \epsilon_{ij} \) is mean 0 error term distributed independently of \( X_j \). The consumer observes the characteristics \( X_j \) of all available products ex-ante, but does not know the function \( m_i(\cdot) \) and must learn about it through search.

The consumer searches through objects sequentially. At each period \( t \), let \( J_{it} \) be the set of objects that have already been searched by consumer \( i \), let \( U_{it} \) be the set of realized utilities of those objects, and let \( \tilde{J}_{it} = J \setminus J_{it} \) be the set of objects that have not been searched. At the beginning of each period, the consumer can choose whether to continue searching or stop and consume one of the previously searched alternatives. If the consumer chooses to search an object \( j \in \tilde{J}_{it} \), she incurs a search cost \( c_{ijt} \), and observes utility \( u_{ij} \). The period then ends, and the consumer starts period \( t + 1 \) with \( J_{it+1} = J_{it} \cup \{j\} \) and \( U_{it+1} = U_{it} \cup \{u_{ij}\} \). If the consumer chooses to stop searching at period \( t \) she consumes the highest utility object from the set of objects that have already been searched, obtaining utility \( \hat{u}_{it} = \max_{j \in U_{it}} u_{ij} \).

The consumer’s state at period \( t \) can therefore be described by the set \( S_{it} \) with typical element \( (j, u_{ij}) \), where \( j \in J_{it} \). That is, the consumer needs to keep track of the set of products that have already been searched, and the observed utility for each of those products. Dropping the \( i \) and \( t \) subscripts, and writing \( S' \) for next period’s state variable, the consumer’s problem can be described by the following Bellman equation.

\[
V(S) = \max \left\{ \hat{u}, \max_{j \in J} \left[ E[V(S')|j, S] - c_j \right] \right\}
\]

(2)

The first term inside the max function is the utility obtained if the consumer stops and purchases the previously searched alternative with the highest utility. The second term inside the max function is the maximum over all products yet to be searched of the expected value of searching a particular product, less the cost of search. The expectation is taken over realizations of the utility of the product to be searched, \( u_{ij} \), with respect to the consumer’s current beliefs about the distribution of that product’s utility. Notice that the consumer can only search each product once, and can only choose to purchase a product that has already been searched.
3.2 Beliefs and Learning

The consumer’s optimal search policy - which products to search at each state and when to stop searching - depends on their beliefs over the utilities of each product. As consumers search, they observe the utilities $u_{ij}$ of products with ex-ante observed characteristics $X_j$. They therefore obtain noisy signals of the unknown function $m_i(X_j)$ which they use to update their beliefs about the utility of products that have not yet been searched. We assume that consumers’ mean utility functions, $m_i(X)$, are random functions drawn from a Gaussian process with prior mean function $\mu_i(X)$ and covariance function $\kappa_i(X, X')$.

$$m_i(X) \sim GP(\mu_i(X), \kappa_i(X, X')) \quad (3)$$

This assumption means that for any finite collection of $J$ products $\{1, ..., J\}$, the vector $(m_i(X_1), ..., m_i(X_J))$ is a multivariate normal random variable with mean $(\mu_i(X_1), ..., \mu_i(X_J))$ and a covariance matrix with $(j,k)$ element $\kappa(X_j, X_k)$. The prior mean function $\mu_i : X \rightarrow \mathbb{R}$ is assumed to be smooth and the covariance function $\kappa_i : X \times X \rightarrow \mathbb{R}$ must be such that the resulting covariance matrix for any $J$ products is symmetric and positive semi-definite. In addition, it is convenient to assume that $\kappa_i(X, X) = \kappa_i(X', X')$ for all $(X, X')$, and that $\kappa_i(X, X') = \tilde{\kappa}_i(\|X - X\|)$, where $\tilde{\kappa}_i(\cdot)$ is a continuous function. That is, the covariance between $m_i(X_j)$ and $m_i(X_k)$ depends only on the distance in characteristic space between products $j$ and $k$. One particular covariance function that satisfies these assumptions is the square exponential covariance function (Rasmussen and Williams, 2006) given by equation 4.

$$\kappa(X, X') = \lambda \exp \left( -\frac{|X - X'|^2}{2\rho^2} \right) \quad (4)$$

Draws of mean utilities from such a Gaussian process are continuous functions over the characteristic space $m_i : X \rightarrow \mathbb{R}$. The parameter $\lambda$ controls the variance of the process. In particular, for any $X_j$, $m_i(X_j) \sim N(\mu_i(X_j), \lambda)$. Notice that as the distance in characteristic space, $|X_j - X_i|$, between two products increases, the covariance of the mean utilities of these products falls. The parameter $\rho$ controls the rate of decay of covariance with distance. Figure 4 illustrates four draws of a Gaussian process with $\mu_i(X) = 0$ where and the square exponential covariance function with $\lambda = 10$, $\rho = 10$, and $X \in [0, 100]$.

Consumers have rational expectations and know the functions $\mu_i(X)$ and $\kappa_i(X, X')$. Each consumer’s prior beliefs over the mean utility function $m_i(X)$ therefore correspond to the true process from which it is drawn. When a consumer chooses to search a product $j$, they observe $u_{ij}$ and update their beliefs over $m_i(X)$ in a Bayesian fashion. We assume that the utility error term $\epsilon_{ij}$ in equation 1 is distributed normal with variance $\sigma^2$, $\epsilon_{ij} \sim N(0, \sigma^2)$, independent across products, $j$, (though not necessarily across consumers, $i$).

Dropping the $i$ subscript, let $\mu(X)$ and $\kappa(X, X')$ be the consumer’s priors. Suppose the consumer searches an object $j$ and observes utility $u_j$. The assumption that $\epsilon_{ij}$ is normally distributed implies that the consumer’s posterior
beliefs over the function \( m(X) \) are given by a Gaussian process with the mean and covariance functions \( \mu'(X) \) and \( \kappa'(X, X') \) given by equations 5 and 6 (Rasmussen and Williams, 2006).

\[
\mu'(X) = \mu(X) + \frac{\kappa(X, X_j)(u_j - \mu(X_j))}{\kappa(X_j, X_j) + \sigma^2} \tag{5}
\]

\[
\kappa'(X, X') = \kappa(X, X') - \frac{\kappa(X, X_j)\kappa(X_j, X')}{\kappa(X_j, X_j) + \sigma^2} \tag{6}
\]

Equation 5 indicates that if the realization \( u_j \) is greater (less) than the prior mean \( \mu(X_j) \), the posterior mean function is revised upwards (downwards) at all points \( X \) in the product space. The magnitude of the change in the mean function at a location \( X \) is proportional to the correlation in the prior Gaussian process between location \( X \) and the searched location \( X_j \). For example, under the square exponential covariance function of equation 4, a realization \( u_j > \mu(X_j) \) would increase the consumer’s mean belief at all points \( X \), with the magnitude of the change being greater for points closer to \( X_j \) in product space. The posterior covariance function, given by equation 6 does not depend on the utility realization \( u_j \), but only on the location in characteristic space \( X_j \) of the searched product. In particular, the variance of the posterior process, \( \kappa'(X, X) \) for any location \( X \) decreases proportionally to the prior covariance between that location and the searched location, \( \kappa(X, X_j) \).

Figure 5 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, \( \mu_i(X) = 0 \) 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 \( m_i(X) \) 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 \( m_i(X) \), especially for products close to \( X_j \) in parameter space.

### 3.3 Consumer Behavior

In theory, consumers should be fully forward looking, evaluating each additional search with an eye to the information that they will gain and could potentially exploit on all future searches. It is however well-known that solving the full dynamic programming problem is computationally intractable unless the number of products is small. For example, one could apply backward induction: for states in which all but one of the products have been searched, the continuation value of searching that remaining product is the expected max of that product’s utility and the highest utility observed so far. The value function at any state can be computed by working backwards from these terminal states. However, this procedure is computationally intensive because the state variables are continuous and unbounded. For example, at a state in which one product has been searched, the realization of that product’s utility could be any real number. Crawford and Shum (2005) deal with this problem in a similar model of learning.
by interpolating the value function between a discrete set of states. This procedure would be difficult using our
data on online search since there are over 300 products. In Crawford and Shum’s (2005) application there are only
5 products.

Instead, algorithms for approximately optimal learning in the machine learning literature on Gaussian processes
typically apply n-period look ahead assumptions, and do not fully solve for the optimal policy function (see for example Osborne et al., 2009). Following this literature, we make the simplifying assumption that consumers only
look one period ahead. That is, at every period consumers compare the utility from stopping now to the utility
from searching one more product and then stopping. Under this assumption, consumers will continue searching if
and only if \( E[\max_{j \in J_{it}} \{ \hat{u}, u_j \} | S] > c \) and when they do so, they will choose the maximizer of \( E[\max_{j \in J_{it}} \{ \hat{u}, u_j \} | S] \).

In the special case of their initial search when their current best option is the outside option (\( \hat{u} = 0 \)) they will just
choose the object for which they have the highest prior utility. As we argue below, this myopic one-period look
ahead policy is actually optimal under certain assumptions.

3.4 Relationship to Other Models

This model of consumer learning is a generalization of two widely used models of consumer choice. First, under
the assumption that there is no correlation across products in the Gaussian process, \( \kappa(X, X') = 0 \) for \( X \neq X' \),
this model is a version of Weitzman’s (1979) model of consumer search. In particular, each product \( j \) has utility
distributed independently according to \( u_{ij} \sim N(\mu_i(X_j), \kappa(X_j, X_j) + \sigma^2) \). When the consumer observes the utility
of product \( j \), she therefore learns nothing about the utility distribution of the other products. As described by
Weitzman (1979), the consumer’s optimal policy is to search products in the order of their reservation utilities, and
stop searching when the best observed utility is greater than the reservation utilities of all un-searched alternatives.

Under the additional assumption that \( \kappa(X, X) = \kappa(X', X') \) for \( X \neq X' \), which is true of, for example, the square
exponential covariance function of equation 4, the distributions of each product’s utility are identical up to mean
shifts. In this case the order of the reservation utilities of alternatives is identical to the order of prior mean utilities
\( \mu_i(X_j) \). The Weitzman optimal policy is therefore to search through the alternatives in the order of \( \mu_i(X_j) \) and
stop searching when the expected gain from searching once more and stopping is less than the search cost. Thus
the optimal policy is a one-period look ahead policy, and so for this class of models the assumption that consumers
are myopic has no bite.

Second, note that with the additional assumption of no search costs, \( c_j = 0 \), consumers will search all products and
purchase the highest utility option. The model then corresponds to a standard random utility model. In particular,
consumer choice probabilities are given by a multinomial probit. The one-period look ahead policy is again optimal
in this case.

\[ ^3\text{In future revisions we hope to explore the implications of this simplifying assumption further, and to attempt to relax it.} \]

\[ ^4\text{If alternative } j \text{ has } u_j \sim f_j(u), \text{ then it has reservation utility } r_j \text{ where } r_j = \int_{r_j}^{\infty} u f_j(u) du + F_j(r_j) r_j. \]
4 Estimation and Identification

4.1 Econometric Specification

In order to take the model to 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:

\[
\mu_i(X_j) = \alpha + X_j \beta_i
\]  
(7)

\[
\beta_{ik} = \exp(\eta_k + \nu_{ik} \sigma_{\beta_k})
\]  
(8)

\[
\nu_{ik} \sim N(0,1)
\]  
(9)

where consumers have individual specific (random) coefficients \( \beta_i \). Recall that \( X_j \) is a \( K \) dimensional vector of product attributes. We assume that each element \( \beta_{ik} \) of the coefficient vector \( \beta_i = (\beta_{i1}, \ldots, \beta_{iK}) \) 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 17. This is similar to the square exponential covariance function of equation 4, 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 \( \rho_k \) that control spatial correlation in utility along the \( K \) dimensions. The parameter \( \lambda \) controls the overall variance level of the prior Gaussian process.

\[
\kappa(X_j, X_l) = \lambda \exp \left( \sum_{k=1}^{K} \frac{(X_{jk} - X_{lk})^2}{2\rho_k^2} \right)
\]  
(10)

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 11 and 12, where \( \mu_c \) and \( \sigma_c \) are parameters, and \( \zeta_{ijt} \) 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.

\[
c_{ijt} = c_i + \zeta_{ijt}
\]  
(11)

\[
\ln(c_i) \sim N(\mu_c, \sigma_c)
\]  
(12)
4.2 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 for this is that a scaling of \( \lambda \) 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.\(^5\)

Thus the parameters to be estimated comprise those determining the distribution of random coefficients in the prior mean, \( \{\eta_k\}_{k=1}^K, \{\sigma_{\beta k}\}_{k=1}^K, \alpha \) \(^6\), 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 \( \theta \) 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.

4.3 Parameter Interpretation and Price Endogeneity

The random coefficients \( \beta_k \) weight the product characteristics in the prior mean. The correct interpretation therefore is that \( \beta_{ik} \) gives the change in the consumer’s prior utility for a one unit change in \( X_k \). In particular, if some observable characteristic \( X_k \) (e.g. price) is positively correlated with some other characteristic that will only be learned after search (e.g. product quality), \( \beta_{ik} \) will account for both the direct (negative) effect of price on prior utility and the indirect (positive) effect of price through its correlation with quality. The standard price endogeneity problem common addressed in industrial organization — that price is correlated with unobserved product attributes — may not arise here, so long as the unobserved product attributes are also unobserved by the consumer at the time of search. Indeed, the possibility of learning product attributes through search is specifically modeled here.

Still, there may be correlation across consumers in the unobserved product attributes (a \( \xi_j \) term). This is not accounted for in the current model, since the Gaussian processes \( m_i(X) \) are drawn iid. We offer an extension with random product effects below to address this concern. The remaining parameters have more straightforward interpretations, capturing the spatial correlation structure and the distribution of search costs respectively.

\(^5\)We do not have a formal non-identification result though, and perhaps the normalization is not strictly necessary.
4.4 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, $\sigma_\beta$, as being identified by the amount of cross-individual variation in searched product attributes relative to within-individual variation. Table 4 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, $\sigma_\beta = 4$. 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, $\alpha$, 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 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.

As a test of identification, we ran our estimation algorithm on simulated data sets for various true parameter values. The estimated parameter values appear to be consistent, and to converge to the true values as sample size is increased. See for example the estimated and true parameters recorded in Table 5.

### 4.5 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 11 with logit errors, the probability of a consumer choosing to search product \( j \in \tilde{J} \) conditional on being at state \( S \), but unconditional on the realizations of the logit cost shocks is given by:

\[
\tilde{P}_i(j|S) = \frac{\exp \left( \mathbb{E}[\max\{\hat{u}, u_j\}]|S| - c \right)}{\exp(\hat{u}) + \sum_{l \in \tilde{J}} \exp \left( \mathbb{E}[\max\{\hat{u}, u_l\}]|S| - c \right)} \tag{13}
\]

Suppose consumer \( i \) searches \( T_i \) times before stopping. Let \( j_{it} \) be the \( t \)th 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, \( S \), was fully observable to the econometrician, the likelihood of the consumer’s search path would then be given by equation 14.

\[
L_i(\{j_{it}\}_{t=0}^{T_i}, \tilde{J}_i, \{S_t\}_{t=0}^{T_i}, \theta, \nu_i, c_i) = \left( \prod_{t=0}^{T_i-1} \tilde{P}_i(j_{it}|S_t) \right) \tilde{P}_i(0|S_{T_i}) \mathbb{1} \left( u_{j_{T_i}} = \hat{u}_{j_{T_i}} \right) \tag{14}
\]

Since the econometrician does not observe the utility draws that enter \( S \), 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 15 corresponding to each utility draw observed by consumer \( i \). Each integral is an expectation with respect to distributions \( F_{ji}(u) = \)
Φ(u_j - m_i(X_j)), where Φ(·) 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 15.

\[
L_i((j_{i,t})_{t=0}^{T_i}, \hat{j}_i | \theta, \nu_i, c_i) = \int \left( \int \ldots \int L_i((j_{i,t})_{t=0}^{T_i}, \hat{j}_i | \{S_t\}_{t=0}^{T_i}, \theta, \nu_i, c_i) dF_{\nu_i}(u_{j_{i,0}}) \ldots dF_{\nu_i}(u_{j_{i,T_i-1}}) \right) dGP(\mu_i, \kappa_i) \quad (15)
\]

Finally, note that the likelihood function in equation 15 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 16 is the resulting likelihood of individual i’s search path and purchase, conditional only on the parameters to be estimated, θ.

\[
L_i((j_{i,t})_{t=0}^{T_i}, \hat{j}_i | \theta) = \int \int \ldots \int L_i((j_{i,t})_{t=0}^{T_i}, \hat{j}_i | \theta, \nu_i, \omega_i) dF_{\nu_i}(\nu_i) dF_c(c_i) \quad (16)
\]

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.

### 4.6 Extensions

**Discrete Characteristics.** In our application, we would like to add brand effects in a parsimonious way. They enter the model in two places. First, we allow brand effects in the prior mean, according to \( \mu_i(X_j) = \alpha + X_j \beta_i + \gamma b(j) \) where b(j) is the brand of product j and the \( \gamma b \) are a set of parameters to be estimated (one for each brand except the first, whose fixed effect is normalized to zero). Second, we amend the covariance function as follows:

\[
\kappa(X_j, X_l) = \exp \left( - \sum_{k=1}^{K} \frac{(X_{jk} - X_{lk})^2}{2\rho_k^2} - \frac{1(b(j) \neq b(l))}{2\rho_b^2} \right) \quad (17)
\]

so that two products with the same brand have the same distance as before, but products with different brands have an additional distance between them that is parameterized by \( \rho_b \).

**Random Product Effects.** To capture the possibility that consumers commonly learn after search that some products are better or worse than others, we add random effects to the model. Specifically, we let \( u_{ij} = m_i(X_j) + \varepsilon_{ij} \) where \( \varepsilon_{ij} = \xi_j + \delta_{ij} \), with \( \delta_{ij} \) drawn from a standard normal distribution, and \( \xi_j \) drawn from a normal distribution with mean zero and variance \( \sigma^2_\xi \). These random effects, which are common across consumers and which consumers
do not observe ex-ante, allow us to rationalize products that are frequently bought (or not bought) conditional on being searched. Products that are rarely bought conditional on being searched are rationalized by a large negative value of $\xi_j$. If consumers update their beliefs spatially, such “surprisingly bad” products should induce subsequent search to take a larger step in attribute space, rationalizing the behaviour documented in Table 3. Note that, from an individual consumer’s perspective, the variance of $\varepsilon_{ij}$ is now $1 + \sigma_\xi^2$. The extra parameter $\sigma_\xi^2$ is identified off of the “excess” correlation across consumers in purchasing decisions, holding the other parameters of the model fixed. Adding them to the likelihood is not hard, though it does considerably complicate estimation, as we now have to integrate out over the random effects in the likelihood function. We use a Gibbs sampling approach in the MCMC algorithm to deal with this.

5 Results

5.1 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. We also omit discrete characteristics such as brand effects, and set $\sigma_\xi = 0$.

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, $\lambda$, and the covariance parameters $\rho_k$ 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 $\rho_k$ 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 cannot reject the hypothesis that $\rho_k > 0$, the data on search paths provides evidence that spatial learning is taking place, and that consumers update their beliefs about un-searched objects as they search. Note that although it is tempting to interpret the $\rho_k$ parameters as indicating that utility draws are more correlated along the price dimension than the display dimension, the relative magnitudes of the $\rho_k$ 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. As discussed in Section 4.3, 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 4, 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.

### 5.2 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 (see for example Ursu, 2016). 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. 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 $\epsilon_{ij}$ for the new product $j$. Although we assumed that $\sigma_\xi = 0$ in the estimated model, in this counterfactual we assume that new product $j$ has $\xi_j = -5$. For this product, we let $\delta_{ij} \sim N(0, \sqrt{0.5})$. We imagine that this particular value of $\gamma_j$ 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 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 $\epsilon_{ij}$, 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 6 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

---

6Note that in this counterfactual, we imagine that all products are viewed on one platform or website, rather than across multiple domains.
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.

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.

6 Conclusion

In this paper we document stylized facts in data on consumer search paths that provide support for the intuitive notion that learning about the payoff from one object provides information on the payoffs from similar objects. We build a model of consumer search in which consumers know some product attributes for each of the available products, but do not know the mapping from these attributes to utility. We introduce the idea of Gaussian process priors on consumer utility, which provide a low dimensional way to capture beliefs about the ex-ante unknown part of utility that are correlated in the space of the known product attributes. Consumer learning is spatial - viewing a product provides more information about other products that are nearby in attribute space than those that are further away. We argue that the parameters of this model are identified from data on consumer search paths and purchases, and estimate the model on data which records the search paths of consumers shopping for digital cameras online.

Finally, we 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


Tables and Figures

Figure 1: Convergence to Chosen Attribute Levels

Notes: 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.
Figure 2: Convergence to Chosen Attribute Levels

Notes: 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.
Notes: 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.
Notes: 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.

Notes: 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.
Figure 6: Substitution from Manipulation of Search Rankings

Notes: 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. Negative 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.

Table 1: Summary Statistics: Products

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>302.13</td>
<td>478.47</td>
<td>16.99</td>
<td>5250.00</td>
</tr>
<tr>
<td>Zoom</td>
<td>6.04</td>
<td>5.60</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>Pixel</td>
<td>10.54</td>
<td>3.13</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>Display</td>
<td>2.67</td>
<td>0.41</td>
<td>1.1</td>
<td>3.5</td>
</tr>
</tbody>
</table>

Notes: 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.
Table 2: Summary Statistics: Search Paths

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Length</td>
<td>5.59</td>
<td>6.52</td>
<td>1</td>
<td>58</td>
</tr>
<tr>
<td>Chosen Product Discovered</td>
<td>0.79</td>
<td>0.29</td>
<td>0.03</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>966</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purchased</td>
<td>211.46</td>
<td>264.85</td>
<td>16.99</td>
<td>2862.03</td>
</tr>
<tr>
<td>Searched</td>
<td>285.45</td>
<td>406.45</td>
<td>16.99</td>
<td>5250</td>
</tr>
<tr>
<td>Zoom</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purchased</td>
<td>6.42</td>
<td>6.02</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>Searched</td>
<td>6.43</td>
<td>5.97</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>Pixel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purchased</td>
<td>12.42</td>
<td>2.16</td>
<td>1.3</td>
<td>21</td>
</tr>
<tr>
<td>Searched</td>
<td>11.96</td>
<td>2.37</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>Display</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purchased</td>
<td>2.82</td>
<td>0.26</td>
<td>1.1</td>
<td>3.5</td>
</tr>
<tr>
<td>Searched</td>
<td>2.78</td>
<td>0.30</td>
<td>1.1</td>
<td>3.5</td>
</tr>
</tbody>
</table>

Notes: 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, but it will not enter the distribution of purchases.
Table 3: Surprisingly Bad Products

| 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 |
| \( SearchPercentile_{it} \) | . -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) | (.094) (.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 | . -299*** -199* -234** |
|                     | (.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 |

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 5%. The additional controls are described in the text. Stars indicate statistical significance as follows: *** p<0.01 ** p<0.05 *p<0.1
Table 4: Identification of $\sigma_\beta$

<table>
<thead>
<tr>
<th>Attribute</th>
<th>$\sigma_\beta = 0$</th>
<th>$\sigma_\beta = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within Individual Variance</td>
<td>1.038</td>
<td>1.023</td>
</tr>
<tr>
<td>Between Individual Variance</td>
<td>0.216</td>
<td>0.388</td>
</tr>
<tr>
<td>$R^2$ of regression on individual fixed effects</td>
<td>0.145</td>
<td>0.195</td>
</tr>
</tbody>
</table>

Notes: 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-dimensional 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.

Table 5: Monte Carlo Results

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>Estimate</th>
<th>SE</th>
<th>True</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_1$</td>
<td>-2</td>
<td>-2.119</td>
<td>0.078</td>
<td>$\rho_1$</td>
<td>15</td>
<td>14.433</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>2</td>
<td>1.977</td>
<td>0.079</td>
<td>$\rho_2$</td>
<td>15</td>
<td>15.885</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-2</td>
<td>-1.507</td>
<td>0.230</td>
<td>$\sigma_\xi$</td>
<td>1</td>
<td>1.212</td>
</tr>
<tr>
<td>$\sigma_{\beta 1}$</td>
<td>1</td>
<td>1.120</td>
<td>0.065</td>
<td>$c$</td>
<td>0</td>
<td>-0.024</td>
</tr>
<tr>
<td>$\sigma_{\beta 2}$</td>
<td>0.5</td>
<td>0.684</td>
<td>0.065</td>
<td>$\sigma_c$</td>
<td>0.8</td>
<td>0.828</td>
</tr>
</tbody>
</table>

Notes: Table records the results of estimating the model on simulated data consisting of 2000 simulated paths. Search is over a 2-dimensional attribute space with 50 products and product attributes randomly drawn from uniform [0, 4] distributions. Random product effects $\xi_j$ are drawn one and held constant over search paths. Estimation uses 7,000 MCMC draws, with the first 2,000 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.

Table 6: Estimated Parameters

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_1$ (log price)</td>
<td>-0.683</td>
<td>0.032</td>
<td>$\rho_1$ (log price)</td>
<td>2.151</td>
</tr>
<tr>
<td>$\eta_2$ (display)</td>
<td>0.292</td>
<td>0.010</td>
<td>$\rho_2$ (display)</td>
<td>1.710</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>5.060</td>
<td>0.270</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\beta 1}$ (log price)</td>
<td>0.599</td>
<td>0.030</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\beta 2}$ (display)</td>
<td>0.133</td>
<td>0.006</td>
<td>$c$</td>
<td>1.627</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\sigma_c$</td>
<td>0.250</td>
</tr>
</tbody>
</table>

Notes: 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.
Table 7: Model Fit

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Simulations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Search Length</td>
<td>5.299</td>
<td>6.466</td>
</tr>
<tr>
<td>Chosen Product Discovered</td>
<td>0.790</td>
<td>0.265</td>
</tr>
<tr>
<td>Within Search Log Price Variance</td>
<td>0.480</td>
<td>0.306</td>
</tr>
<tr>
<td>Within Search Display Variance</td>
<td>0.236</td>
<td>0.148</td>
</tr>
<tr>
<td>Distance of First Search to Purchase in Log Price</td>
<td>0.359</td>
<td>0.545</td>
</tr>
<tr>
<td>Distance of First Search to Purchase in Display</td>
<td>0.165</td>
<td>0.258</td>
</tr>
</tbody>
</table>

Notes: 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 2, at which the product eventually chosen is first viewed.

Table 8: Manipulation of Search Rankings

<table>
<thead>
<tr>
<th></th>
<th>Free Search</th>
<th>Forced Search</th>
<th>Counterfactual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Length</td>
<td>5.299</td>
<td>5.910</td>
<td></td>
</tr>
<tr>
<td>Outside Option Share</td>
<td>12%</td>
<td>17%</td>
<td></td>
</tr>
<tr>
<td>Mean Consumption Utility: Learning Model</td>
<td>3.956</td>
<td>3.751</td>
<td></td>
</tr>
<tr>
<td>Mean Consumption Utility: No Learning Model</td>
<td>4.232</td>
<td>4.235</td>
<td></td>
</tr>
</tbody>
</table>

Notes: 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.