Microsoft Research

Each year Microsoft Research hosts hundreds of influential speakers from around the world including leading scientists, renowned experts in technology, book authors, and leading academics, and makes videos of these lectures freely available.

2016 © Microsoft Corporation. All rights reserved.
You can lead a horse to water: spatial learning and path dependence in consumer search

Charles Hodgson
Greg Lewis
Goal

To introduce an empirically tractable framework for modeling consumer search with learning
Search

Search = costly information or opportunity acquisition, with the goal of solving a decision problem

Examples:

Good search (1-sided) (e.g. cameras)
House search (1-sided)
Dating (2-sided)
Job search (2-sided)
Search = costly information or opportunity acquisition, with the goal of solving a decision problem.

Examples:

Good search (1-sided) (e.g. cameras)
House search (1-sided)
Dating (2-sided)
Job search (2-sided)

Why study search?
Search

Search = costly information or opportunity acquisition, with the goal of solving a decision problem

Examples:

Good search (1-sided) (e.g. cameras)
House search (1-sided)
Dating (2-sided)
Job search (2-sided)
Why study search?
“In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.”
— Herbert A. Simon

(Designing organizations for an information-rich world, 1971)
Figure 3: Click through rate (CTR) by results position and search impression type

Ursu (2017)
“In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.”
— Herbert A. Simon

(Designing organizations for an information-rich world, 1971)
Figure 3: Click through rate (CTR) by results position and search impression type

Ursu (2017)
Expedia’s Ranking

Random Ranking

Ursu (2017)
Most sequential search models are urn models.

Opportunities look identical before search; and only look different after search.

This is a poor description of many search environments.
Expedia’s Ranking

Random Ranking

Ursu (2017)
Most sequential search models are urn models.

Opportunities look identical before search; and only look different after search.

This is a poor description of many search environments.
Ursu (2017)
Most sequential search models are urn models.

Opportunities look identical before search; and only look different after search

This is a poor description of many search environments.
Formal model (e.g. McCall 1970)

- Payoffs $u_i \sim F$
- As many samples can be drawn as you’d like
- Each sample costs $c$
- Solution: a stopping rule – stop whenever you get $u_i > \bar{u}$
  (“good enough” / “satisficing”)
Challenge 1: Observables
Formal model (Weitzman 1979)

- Finite set of options (boxes)
- Each has its own payoff distribution $F_i$
- Actual payoff $u_i \sim F_i$, iid across boxes
- Each box has a cost to open $c_i$
- Solution: give each box a score equal to the current payoff that would make you indifferent between sampling that box and stopping with that payoff
- Open them from top score to bottom, stopping if the current payoff ever exceeds the next score
Formal model (Weitzman 1979)

- Finite set of options (boxes)
- Each has its own payoff distribution $F_i$
- Actual payoff $u_i \sim F_i$, iid across boxes
- Each box has a cost to open $c_i$
- Solution: give each box a score equal to the current payoff that would make you indifferent between sampling that box and stopping with that payoff
- Open them from top score to bottom, stopping if the current payoff ever exceeds the next score
Challenge 1: Observables
Challenge 2: Learning
What is learning?

Learning = updating beliefs about other options based on the past samples

How should you update?

Depends on the correlation structure...but where does that come from?

Our answer: people update more for “close-by” objects, where “close-by” is defined by a feature space
NOPE. NOT TODAY.

Challenge 3: Revisits, Tabbing, Duration
What is learning?

Learning = updating beliefs about other options based on the past samples

How should you update?

Depends on the correlation structure...but where does that come from?

Our answer: people update more for “close-by” objects, where “close-by” is defined by a feature space
NOPE. NOT TODAY.

Challenge 3: Revisits, Tabbing, Duration
Contributions

1) Introduce a tractable model of “spatial learning” : sampling one product updates beliefs about nearby products in attribute space

2) Show the possibility of path dependence and belief manipulation

3) Show evidence of spatial learning in camera search data

4) Fit a structural model to the camera data, and illustrate the impact of a re-ordering of the products on welfare
A Motivating Example

- 3 Products: A, B, C
- $u_j = q_j - p_j$, $j = A, B, C$
- Price observable
- Quality only observable after clicking on the product, at search cost c.
- Quality is drawn from $q \sim N(\mu, \Sigma)$, $\mu > 1$

What’s the optimal search strategy?
Independence
A Motivating Example

• 3 Products: A, B, C
• \( u_j = q_j - p_j \), \( j = A, B, C \)
• Price observable
• Quality only observable after clicking on the product, at search cost c.
• Quality is drawn from \( q \sim N(\mu, \Sigma), \mu > 1 \)

What’s the optimal search strategy?
Independence
Positive Correlation
What a bad experience does

What if I could make the consumer sample B first?
Purchase Manipulation

<table>
<thead>
<tr>
<th>Starting Product</th>
<th>$c \in [0, 0.05]$</th>
<th>$c \in (0.05, 0.78]$</th>
<th>$c \in (0.78, 0.91]$</th>
<th>$c \in (0.91, \infty]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>

Strategies to get A purchased:

- with low search costs, nothing works
- with medium search costs, put B first
- with high search costs, put A first
What a bad experience does

What if I could make the consumer sample B first?
## Purchase Manipulation

<table>
<thead>
<tr>
<th>Starting Product</th>
<th>$c \in [0, 0.05]$</th>
<th>$c \in (0.05, 0.78]$</th>
<th>$c \in (0.78, 0.91]$</th>
<th>$c \in (0.91, \infty]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>

**Strategies to get A purchased:**

- with low search costs, nothing works
- with medium search costs, put B first
- with high search costs, put A first
What a bad experience does

What if I could make the consumer sample B first?
Purchase Manipulation

<table>
<thead>
<tr>
<th>Starting Product</th>
<th>$c \in [0, 0.05]$</th>
<th>$c \in (0.05, 0.78]$</th>
<th>$c \in (0.78, 0.91]$</th>
<th>$c \in (0.91, \infty]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>

Strategies to get A purchased:

- with low search costs, nothing works
- with medium search costs, put B first
- with high search costs, put A first
A General Model

- Consumers have utility drawn as:

  \[ u_j(X) = m(X) + \varepsilon_j \]

- \( X \) is the characteristic space, observables at time of search
- \( m(X) \) is the predictable part of utility, \( \varepsilon \) is the idiosyncratic part (iid Gaussian draws)
- Both \( m(X) \) and \( \varepsilon \) are unknown, but consumers learn about \( m(X) \)
- Interpretation 1: all common characteristics are observable, but consumers don’t know how to value them, they learn by clicking/reading
- Interpretation 2: consumers know how to value all characteristics, but some of them are unobserved, and they’re learning the mapping
# Purchase Manipulation

<table>
<thead>
<tr>
<th>Starting Product</th>
<th>$c \in [0, 0.05]$</th>
<th>$c \in (0.05, 0.78]$</th>
<th>$c \in (0.78, 0.91]$</th>
<th>$c \in (0.91, \infty]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>

**Strategies to get A purchased:**

- with low search costs, nothing works
- with medium search costs, put B first
- with high search costs, put A first
A General Model

• Consumers have utility drawn as:

$$u_j(X) = m(X) + \varepsilon_j$$

• $X$ is the characteristic space, observables at time of search
• $m(X)$ is the predictable part of utility, $\varepsilon$ is the idiosyncratic part (iid Gaussian draws)
• Both $m(X)$ and $\varepsilon$ are unknown, but consumers learn about $m(X)$
• Interpretation 1: all common characteristics are observable, but consumers don’t know how to value them, they learn by clicking/reading
• Interpretation 2: consumers know how to value all characteristics, but some of them are unobserved, and they’re learning the mapping
A General Model

- The function \( m(X) \) is drawn from a Gaussian process

\[
m(X) \sim (\mu(X), \kappa(X, X'))
\]

- Mean function: what’s the expected value at a point?
- Covariance (kernel) function: what’s the relationship between the realizations at different points?
- Consumers update over time.
- Mean depends on past realizations; covariance depends only on where was searched.
\[ \kappa(X, X') = \lambda \exp \left( \frac{-|X - X'|^2}{2\rho^2} \right) \]
Optimal search

The dynamic search problem has a Bellman equation characterization, but that’s about it.

Instead, we consider 1-period look-ahead search:
- consumers think about either continuing and purchasing the best product available, or making one more search then stopping.

Myopic = ignores choice for information acquisition, highly speculative options.

Optimal in our setting under independence
\[ \kappa(X, X') = \lambda \exp\left( -\frac{|X - X'|^2}{2\rho^2} \right) \]
Optimal search

The dynamic search problem has a Bellman equation characterization, but that’s about it.

Instead, we consider 1-period look-ahead search:
- consumers think about either continuing and purchasing the best product available, or making one more search then stopping.

Myopic = ignores choice for information acquisition, highly speculative options.

Optimal in our setting under independence
A General Model

- The function $m(X)$ is drawn from a Gaussian process

$$m(X) \sim (\mu(X), \kappa(X, X'))$$

- Mean function: what’s the expected value at a point?
- Covariance (kernel) function: what’s the relationship between the realizations at different points?
- Consumers update over time.
- Mean depends on past realizations; covariance depends only on where was searched.
A General Model

- The function $m(X)$ is drawn from a Gaussian process

\[ m(X) \sim (\mu(X), \kappa(X, X')) \]

- Mean function: what’s the expected value at a point?
- Covariance (kernel) function: what’s the relationship between the realizations at different points?
- Consumers update over time.
- Mean depends on past realizations; covariance depends only on where was searched.
A General Model

• Consumers have utility drawn as:

\[ u_j(X) = m(X) + \varepsilon_j \]

• \( X \) is the characteristic space, observables \textit{at time of search}
• \( m(X) \) is the predictable part of utility, \( \varepsilon \) is the idiosyncratic part (iid Gaussian draws)
• Both \( m(X) \) and \( \varepsilon \) are unknown, but consumers learn about \( m(X) \)
• Interpretation 1: all common characteristics are observable, but consumers don’t know how to value them, they learn by clicking/reading
• Interpretation 2: consumers know how to value all characteristics, but some of them are unobserved, and they’re learning the mapping
A General Model

• Consumers have utility drawn as:

\[ u_j(X) = m(X) + \varepsilon_j \]

• \( X \) is the characteristic space, observables *at time of search*
• \( m(X) \) is the predictable part of utility, \( \varepsilon \) is the idiosyncratic part (iid Gaussian draws)
• Both \( m(X) \) and \( \varepsilon \) are unknown, but consumers learn about \( m(X) \)
• Interpretation 1: all common characteristics are observable, but consumers don’t know how to value them, they learn by clicking/reading
• Interpretation 2: consumers know how to value all characteristics, but some of them are unobserved, and they’re learning the mapping
A General Model

• Consumers have utility drawn as:

\[ u_j(X) = m(X) + \varepsilon_j \]

• \( X \) is the characteristic space, observables at time of search

• \( m(X) \) is the predictable part of utility, \( \varepsilon \) is the idiosyncratic part (iid Gaussian draws)

• Both \( m(X) \) and \( \varepsilon \) are unknown, but consumers learn about \( m(X) \)

• Interpretation 1: all common characteristics are observable, but consumers don’t know how to value them, they learn by clicking/reading

• Interpretation 2: consumers know how to value all characteristics, but some of them are unobserved, and they’re learning the mapping
A General Model

- The function $m(X)$ is drawn from a Gaussian process

$$m(X) \sim (\mu(X), \kappa(X, X'))$$

- Mean function: what’s the expected value at a point?
- Covariance (kernel) function: what’s the relationship between the realizations at different points?
- Consumers update over time.
- Mean depends on past realizations; covariance depends only on where was searched.
$$\kappa(X, X') = \lambda \exp \left( -\frac{|X - X'|^2}{2\rho^2} \right)$$
Optimal search

The dynamic search problem has a Bellman equation characterization, but that’s about it.

Instead, we consider 1-period look-ahead search:
- consumers think about either continuing and purchasing the best product available, or making one more search then stopping.

Myopic = ignores choice for information acquisition, highly speculative options.

Optimal in our setting under independence
Optimal search

The dynamic search problem has a Bellman equation characterization, but that’s about it.

Instead, we consider 1-period look-ahead search:
- consumers think about either continuing and purchasing the best product available, or making one more search then stopping.

Myopic = ignores choice for information acquisition, highly speculative options.

Optimal in our setting under independence
\[ \kappa(X, X') = \lambda \exp \left( \frac{-|X - X'|^2}{2\rho^2} \right) \]
Optimal search

The dynamic search problem has a Bellman equation characterization, but that’s about it.

Instead, we consider 1-period look-ahead search:
- consumers think about either continuing and purchasing the best product available, or making one more search then stopping.

Myopic = ignores choice for information acquisition, highly speculative options.

Optimal in our setting under independence
The one-period look ahead policy scores the available options based on their expected marginal contribution over the current best option \( \hat{u} \). Define \( \alpha_j = (\hat{u} - \mu(x_j))/\sqrt{\kappa(x_j, x_j) + \sigma^2} \), the current best option normalized by the mean and variance of item \( j \). Then we score option \( j \) according to:

\[
    z_j = \Phi(\alpha_j)\hat{u} + (1 - \Phi(\alpha_j))\mu_j + \phi(\alpha_j)\sqrt{\kappa(x_j, x_j) + \sigma^2} - c_j
\]

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.
Interpretation

The search index is increasing in the value of the current best option, the mean and variance of the option being considered, and decreasing in search cost.

Consumers will opt for options with high mean and high variance.

Consumers will be more likely to go for risky options when their current option is good. Risky = bigger left tail, downside risk, second-order stochastically dominated.
The Next Step

What happens after a consumer searches product k? How do they adjust future behavior?

A good experience (high payoff) at location k increases the posterior mean at places with high covariance with k. Whether this update is pivotal will depend on the current best option.
The Next Step

What happens after a consumer searches product k? How do they adjust future behavior?

A good experience (high payoff) at location k increases the posterior mean at places with high covariance with k. Whether this update is pivotal will depend on the current best option.
The Next Step

Implication 1: After a bad experience with a product, a consumer should jump to somewhere either far away or known/safe.

Implication 2: A platform can steer consumers away from a part of the characteristic space by ranking bad products from that area highly.
Data

- From Comscore, via Bronnenberg, Kim, Mela (2016)
- Sequence of URLs, across websites
- We see the sequence of product pages loaded, as well as the product eventually purchased (if purchased)
- 967 panelists, August-Dec 2010
The Next Step

Implication 1: After a bad experience with a product, a consumer should jump to somewhere either far away or known/safe

Implication 2: A platform can steer consumers away from a part of the characteristic space by ranking bad products from that area highly
Data

• From Comscore, via Bronnenberg, Kim, Mela (2016)
• Sequence of URLs, across websites
• We see the sequence of product pages loaded, as well as the product eventually purchased (if purchased)
• 967 panelists, August-Dec 2010
<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><strong>N</strong></td>
<td>966</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Price</strong></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><strong>Zoom</strong></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><strong>Pixel</strong></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><strong>Display</strong></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>
Zooming in (BKM)
Zooming in (BKM)
Smaller steps (us)
A Test of Spatial Learning

Theory: bad experience $\rightarrow$ move to another part of the characteristic space

Define “bad product” = viewed by at least 10 people, purchased by less than 5%

Regress step size on the previous product being bad, adding controls for search percentile (how far along), product density (# of products within 1 SD), consumer FEs
Smaller steps (us)
A Test of Spatial Learning

Theory: bad experience $\rightarrow$ move to another part of the characteristic space

Define “bad product” = viewed by at least 10 people, purchased by less than 5%

Regress step size on the previous product being bad, adding controls for search percentile (how far along), product density (# of products within 1 SD), consumer FEs
| Dependent Variable: | \( |Price_{it} - Price_{it-1}| \) |
|---------------------|---------------------------------|
|                    | \( BadProduct_{it-1} \)        | \( SearchPercentile_{it} \) | \( Constant \) |
|                    | 103.109*** 100.824*** 68.291*** | -93.074*** -62.885*** -54.060*** | 170.110*** 225.209*** 980.316*** |

**Density Controls**
- No No Yes Yes

**Individual FE**
- No No No Yes

**N**
- 6526 6526 6526 6526
A Test of Spatial Learning

Theory: bad experience -> move to another part of the characteristic space

Define “bad product” = viewed by at least 10 people, purchased by less than 5%

Regress step size on the previous product being bad, adding controls for search percentile (how far along), product density (# of products within 1 SD), consumer FEs
| Dependent Variable: | \( |Price_{it} - Price_{it-1}| \) |
|---------------------|---------------------------------|
| \( \text{BadProduct}_{it-1} \) | 103.109*** 100.824*** 68.291*** 62.003** |
| \( \text{SearchPercentile}_{it} \) | \( \cdot \) -93.074*** -62.885*** -54.060*** |
|                     | \( \cdot \) (17.209) (15.630) (16.402) |
| \( \text{Constant} \) | 170.110*** 225.209*** 980.316*** 921.834*** |
|                     | (4.964) (11.328) (27.681) (34.994) |
| Density Controls    | No  No  Yes  Yes |
| Individual FE       | No  No  No  Yes |
| \( N \)             | 6526 6526 6526 6526 |
A Structural Model

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

(7)

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

(8)

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

(9)

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

(10)

\[ \ln(c_i) \sim N(\mu_c, \sigma_c) \]  

(12)
| Dependent Variable: | $|Price_{it} - Price_{it-1}|$ |
|---------------------|-------------------------------|
| $BadProduct_{it-1}$ | 103.109*** 100.824*** 68.291*** 62.003** |
| $SearchPercentile_{it}$ | . -93.074*** -62.885*** -54.060*** |
|                     | .  (17.209)  (15.630)  (16.402) |
| Constant            | 170.110*** 225.209*** 980.316*** 921.834*** |
|                     | (4.964)  (11.328)  (27.681)  (34.994) |
| Density Controls    | No  No  Yes  Yes |
| Individual FE       | No  No  No  Yes |
| $N$                 | 6526 6526 6526 6526 |
A Structural Model

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

(7)

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

(8)

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

(9)

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

(10)

\[ \ln(c_i) \sim N(\mu_c, \sigma_c) \]  

(12)
| Dependent Variable: | $|Price_{it} - Price_{it-1}|$ |
|---------------------|-----------------------------|
|                     | $BadProduct_{it-1}$ | $103.109^{***}$ | $100.824^{***}$ | $68.291^{***}$ | $62.003^{**}$ |
|                     | $SearchPercentile_{it}$ | $-93.074^{***}$ | $-62.885^{***}$ | $-54.060^{***}$ |
|                     | .                 | (17.209)        | (15.630)        | (16.402)        |
|                     | $Constant$        | $170.110^{***}$ | $225.209^{***}$ | $980.316^{***}$ | $921.834^{***}$ |
|                     | (4.964)           | (11.328)        | (27.681)        | (34.994)        |
| Density Controls    | No                | No              | Yes             | Yes             |
| Individual FE       | No                | No              | No              | Yes             |
| $N$                 | 6526              | 6526            | 6526            | 6526            |
A Structural Model

\[ \mu_i(X_j) = \alpha + X_j \beta_i \]  \hspace{1cm} (7)

\[ \beta_{ik} = \exp(\eta_k + \nu_{ik} \sigma_{\beta k}) \]  \hspace{1cm} (8)

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

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

\[ \ln(c_i) \sim N(\mu_c, \sigma_c) \]  \hspace{1cm} (12)
# Model Fit

<table>
<thead>
<tr>
<th></th>
<th>Data Mean</th>
<th>Data SD</th>
<th>Simulations Mean</th>
<th>Simulations SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Search Length</strong></td>
<td>5.299</td>
<td>6.466</td>
<td>5.436</td>
<td>7.187</td>
</tr>
<tr>
<td>Chosen Product Discovered</td>
<td>0.790</td>
<td>0.265</td>
<td>0.718</td>
<td>0.295</td>
</tr>
<tr>
<td>Within Search Log Price Variance</td>
<td>0.480</td>
<td>0.306</td>
<td>0.886</td>
<td>0.418</td>
</tr>
<tr>
<td>Within Search Display Variance</td>
<td>0.236</td>
<td>0.148</td>
<td>0.382</td>
<td>0.224</td>
</tr>
<tr>
<td>Distance of First Search to Purchase in Log Price</td>
<td>0.359</td>
<td>0.545</td>
<td>0.633</td>
<td>0.838</td>
</tr>
<tr>
<td>Distance of First Search to Purchase in Display</td>
<td>0.165</td>
<td>0.258</td>
<td>0.288</td>
<td>0.434</td>
</tr>
</tbody>
</table>
What a bad experience does

Bad for close by products, good for far away

(1000 simulated search paths)
## Model Fit

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th></th>
<th>Simulations</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Search Length</td>
<td>5.299</td>
<td>6.466</td>
<td>5.436</td>
<td>7.187</td>
</tr>
<tr>
<td>Chosen Product Discovered</td>
<td>0.790</td>
<td>0.265</td>
<td>0.718</td>
<td>0.295</td>
</tr>
<tr>
<td>Within Search Log Price Variance</td>
<td>0.480</td>
<td>0.306</td>
<td>0.886</td>
<td>0.418</td>
</tr>
<tr>
<td>Within Search Display Variance</td>
<td>0.236</td>
<td>0.148</td>
<td>0.382</td>
<td>0.224</td>
</tr>
<tr>
<td>Distance of First Search to Purchase in Log Price</td>
<td>0.359</td>
<td>0.545</td>
<td>0.633</td>
<td>0.838</td>
</tr>
<tr>
<td>Distance of First Search to Purchase in Display</td>
<td>0.165</td>
<td>0.258</td>
<td>0.288</td>
<td>0.434</td>
</tr>
</tbody>
</table>
What a bad experience does

Bad for close by products, good for far away

(1000 simulated search paths)
## Model Fit

<table>
<thead>
<tr>
<th>Metric</th>
<th>Data Mean</th>
<th>Data SD</th>
<th>Simulations Mean</th>
<th>Simulations SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Length</td>
<td>5.299</td>
<td>6.466</td>
<td>5.436</td>
<td>7.187</td>
</tr>
<tr>
<td>Chosen Product Discovered</td>
<td>0.790</td>
<td>0.265</td>
<td>0.718</td>
<td>0.295</td>
</tr>
<tr>
<td>Within Search Log Price Variance</td>
<td>0.480</td>
<td>0.306</td>
<td>0.886</td>
<td>0.418</td>
</tr>
<tr>
<td>Within Search Display Variance</td>
<td>0.236</td>
<td>0.148</td>
<td>0.382</td>
<td>0.224</td>
</tr>
<tr>
<td>Distance of First Search to Purchase in Log Price</td>
<td>0.359</td>
<td>0.545</td>
<td>0.633</td>
<td>0.838</td>
</tr>
<tr>
<td>Distance of First Search to Purchase in Display</td>
<td>0.165</td>
<td>0.258</td>
<td>0.288</td>
<td>0.434</td>
</tr>
</tbody>
</table>
What a bad experience does

Bad for close by products, good for far away

(1000 simulated search paths)
## Counterfactual Results

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

Bad for close by products, good for far away

(1000 simulated search paths)
# Counterfactual Results

<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>