LIU (CATHY) YANG, OLIVIER TOUBIA, and MARTIJN G. DE JONG*

It is becoming increasingly easier for researchers and practitioners to collect eye-tracking data during online preference measurement tasks. The authors develop a dynamic discrete choice model of information search and choice under bounded rationality, which they calibrate using a combination of eye-tracking and choice data. Their model extends Gabaix et al.'s (2006) directed cognition model by capturing fatigue, proximity effects, and imperfect memory encoding and by estimating individual-level parameters and partworths within a likelihood-based hierarchical Bayesian framework. The authors show that modeling eye movements as the outcome of forward-looking utility maximization improves out-of-sample predictions, enables researchers and practitioners to use shorter questionnaires, and allows better discrimination between attributes.

Keywords: preference measurement, eye tracking, dynamic discrete choice models

Online Supplement: http://dx.doi.org/10.1509/jmr.13.0288

A Bounded Rationality Model of Information Search and Choice in Preference Measurement

Choice experiments are used routinely in the fields of marketing, economics, and psychology. One common example is choice-based conjoint (CBC) analysis. An implicit assumption made in standard choice-based preference measurement models is that respondents are fully rational and thus can systematically process all the choice-relevant information (i.e., attribute levels of all alternatives) and choose the alternative that provides the greatest utility. However, the bounded rationality literature (Simon 1955) has found that this assumption is not necessarily valid, and consumers have been shown to balance the utility of the option they choose with the (cognitive) utility derived from the choice process itself (e.g., Payne, Bettman, and Johnson 1988, 1992, 1993).[1]

Trading off the costs of processing information with the benefits from the choice leads to some choice-relevant information not being processed at all. This phenomenon has long been recognized in the marketing literature (e.g., Hagerty and Aaker 1984; Hauser, Urban, and Weinberg 1993; Meyer 1982) and has been documented using eye-tracking evidence (Shi, Wedel, and Pieters 2013; Stüttgen, Boatwright, and Monroe 2012; Toubia et al. 2012). Recent models have endogenized the information acquisition process (Shi, Wedel, and Pieters 2013; Stüttgen, Boatwright, and Monroe 2012), but not in a way that explicitly captures the dynamic trade-off between the effort spent acquiring

*Liu (Cathy) Yang is Assistant Professor, HEC Paris (e-mail: yang@hec.fr). Olivier Toubia is Glaubinger Professor of Business, Columbia Business School, Columbia University (e-mail: ot2107@columbia.edu). Martijn G. de Jong is Professor of Marketing Research and Tinbergen Research Fellow, Erasmus School of Economics (e-mail: mgdejong@ese.eur.nl). The authors are grateful to the *JMR* review team for constructive comments on this article. This article has greatly benefited from Joel Huber's feedback. Michel Wedel served as associate editor for this article.

[1]In doing so, they may revert to noncompensatory decision rules—for example, disjunctive, conjunctive (Gilbride and Allenby 2004, 2006; Jedidi and Kohli 2005), lexicographic, and elimination by aspects (Johnson, Meyer, and Ghose 1989; Payne, Bettman, and Johnson 1988; Tversky, Sattath, and Slovic 1988; Yee et al. 2007). From a modeling perspective, we leverage the fact that noncompensatory decision rules are nested within additive utility models, and we do not model them directly (Jedidi and Kohli 2005, 2008; Yee et al. 2007).

ISSN: 0022-2437 (print), 1547-7193 (electronic)

information and the benefits of making better-informed decisions.

Our article is an attempt to close this gap. We develop a dynamic discrete choice model of information processing and choice under bounded rationality, which we calibrate using a combination of eye-tracking and choice data. To the extent that (1) consumers are strategic in their information acquisition process and (2) information acquisition is motivated by utility maximization, the information acquisition process should contain valuable information about consumers' preferences. In our model, one eye fixation corresponds to one time period. The information acquisition process is driven at each period by exogenous factors (e.g., display layout) as well as endogenous factors (e.g., the information acquired up to that point). The information acquired in each period influences the evaluation of the choice alternatives, which will in turn affect how information is acquired in the following periods. This feedback loop, combined with the forward-looking nature of our model, allows information acquisition in each period to be influenced by how it will affect the future evaluation of alternatives. We show that complementing choice data with eye-tracking data and modeling eye movements as outcomes of forward-looking utility maximization can improve out-of-sample performance, enable practitioners and researchers to use shorter questionnaires, and allow better discrimination between attributes.

Although we collected our eye-tracking data in a dedicated lab, commercial solutions are available, such as EyeTrackShop (www.eyetrackshop.com) and YouEye (www.youeye.com), that allow for collection of eye-tracking data in an online environment using the consumer's webcam. Therefore, we believe that the model developed in this article and the data on which it relies will be widely accessible in the near future and that market researchers will be able to collect eye-tracking data systematically to augment traditional choice data.

The rest of the article is organized as follows. In the next section, we review some relevant prior literature. Then, we present our model, describe our data, report the estimation results, and discuss our conclusions.

PRIOR LITERATURE

Our model bridges the literature on dynamic discrete choices and eye tracking. Before reviewing these literature streams, we briefly introduce the context and type of data considered in our model. We consider a consumer who makes a series of choices in which alternatives are described by attributes that may have several levels. We assume that the choice-relevant information is presented to the consumer in a matrix such as that shown in Figure 1, with one column per alternative and one row per attribute (alternative formats could be modeled as well, as in Shi, Wedel, and Pieters [2013]). We also assume that we observe, for each choice question, a series of eye fixations that end when the consumer chooses one of the alternatives. On each search opportunity, the consumer makes a choice between acquiring some choice-relevant information (by visiting a cell in the matrix) or ending the search and choosing one of the alternatives on the basis of the information acquired up to that point. In the latter case, the consumer moves on to the next choice question.

Dynamic Models of Search

The decisions made by consumers in the process of acquiring choice-relevant information and choosing one alternative are an example of a typical dynamic choice setting, in which each decision may affect the utility offered by various future possible decisions. For example, acquiring a new piece of information on one alternative may change the identity of the alternative in the choice set with the highest expected utility. Such choice problems can be modeled using dynamic discrete choice models (e.g., Ching et al. 2012; Chintagunta, Goettler, and Kim 2012; Dubé, Hitsch, and Jindal 2012; Hartmann and Nair 2010; Huang, Khwaja, and Sudhir 2012; Misra and Nair 2011; Rust 1987; Toubia and Stephen 2013; Yao and Mela 2011). However, the standard approach to dynamic discrete choice modeling poses at least three challenges in our case. First, the state space is likely to be too large to allow estimation of a traditional dynamic discrete choice model using the tools and computers available today. For example, suppose there are four alternatives per choice question described by six attributes. In this typical scenario, simply keeping track of which cells of the matrix the consumer visited would require 2^{24} possible states. Second, such an approach would not fit well with the assumption that consumers trade off decision accuracy and decision cost. In particular, when it is assumed that processing information and making decisions are potentially costly, search models that require solving dynamic programs suffer from the "infinite regress problem"—that is, agents should optimize how they will optimize their decisions and optimize how they will optimize the way they optimize their decisions, and so on (Gabaix et al. 2006). Third, a standard dynamic discrete choice model is not likely to be ecologically valid in our case because it would be challenging for the human brain to implement within the time frame considered in our data, in which eye fixations take place every 99.42 milliseconds, on average. Several researchers have argued that models based on dynamic programming, while normative, should be adjusted to capture the behavior of boundedly rational consumers (e.g., Assunção and Meyer 1993; Hutchinson and Meyer 1994). For example, Camerer, Ho, and Chong (2004) find that, given the constraints imposed by working memory, models that are forward looking by only one or two steps fit data better than fully forward-looking models.

In light of these issues, we base our model on the directed cognition (DC) model proposed and validated by Gabaix et al. (2006). According to this model, on each search occasion t, the participant chooses as if this search occasion were the last one in that question. In other words, if the consumer decides to acquire some information, he or she does so as if (s)he would be making a choice immediately after acquiring this new piece of information.

The DC model offers several benefits in addition to being computationally tractable for complex search problems such as ours. First, although the DC model is not fully forward looking, neither is it myopic; it does capture the basic trade-off in search problems (i.e., search more and choose later using more information vs. choose now using the cur-

Figure 1
SCREENSHOT FROM THE FIRST QUESTION IN THE MAIN TASK

PART I - QUESTION 1

Please indicate your favorite product from the set below.
One out of 100 respondents will be selected as a winner and will receive 800 euros, which will be used to purchase a laptop automatically.
If you are selected as a winner, with 50% probability you will receive your preferred laptop from Part II. With probability 50%, you will receive your preferred laptop from one randomly selected question from Part I. All questions from Part I are equally likely to be selected.
In all cases, you will receive both the laptop and the difference between 800 euros and its price.

	A	B	C	D
Processor speed	2.7 Ghz	2.7 Ghz	1.6 Ghz	3.2 Ghz
Screen size	40 cm	43 cm	40 cm	35.6 cm
Hard drive	160 GB	500 GB	320 GB	320 GB
Dell support	3 years	2 years	2 years	4 years
McAfee subscription	2 years	3 years	2 years	2 years
Price	500 euro	650 euro	500 euro	800 euro
	○	○	○	○

submit

rently available information). Second, there is evidence that this model describes the actual behavior of human agents better than traditional search models based on optimal solutions to dynamic programs. Using a simple experimental setting, Gabaix et al. (2006) show that the DC model predicts behavior better than a search model based on the optimal strategy, which in that case was available in closed form, using the Gittins–Weitzman index (Gittins 1979; Weitzman 1979). In another experiment, the authors show that the model predicts behavior well in a more complex setup with some similarities to CBC analysis. In particular, Gabaix et al. used a setting in which choices were presented in a matrix format similar to Figure 1. However, each cell contained monetary payoffs, and the value of each alternative was the sum of the amounts in the corresponding cells (i.e., participants received the monetary value of the chosen alternative). The authors tested the DC model using a MouseLab paradigm (see Payne, Bettman, and Johnson 1993) in which most information was hidden and participants could "open" only one cell at a time.

Although the DC model provides us with a framework that informs our modeling efforts, our model differs significantly from that used by Gabaix et al. (2006). We compare our implementation of the DC model with theirs in the "Comparison with Gabaix et al. (2006)" subsection after describing our model in more detail.

We also note that Gabaix et al.'s (2006) DC model is related to previous studies in the marketing literature. For

example, Hagerty and Aaker (1984) consider a similar context in which information is presented to consumers in a matrix form. In their model, at each search opportunity, consumers evaluate the expected gain from visiting each cell in the matrix. That gain is linked to the probability that visiting a piece of information will change the identity of the option that provides the greatest expected utility. Like Gabaix et al., Hagerty and Aaker assume that consumers select the cell in the matrix that will maximize the expected gain in utility in the next period. For a related model, see Meyer (1982).

Eye-Tracking Research in Marketing

Eye-tracking data are composed of fixations and saccades (Wedel and Pieters 2000). Fixations represent the time periods in which participants fix their eyesight on a specific location; saccades represent eye movements between two fixations. As a way to directly measure attention and involvement, eye-tracking studies have been conducted in numerous marketing settings, including branding (Pieters and Warlop 1999; Van der Lans, Pieters, and Wedel 2008a), advertising (Pieters, Rosbergen, and Wedel 1999; Pieters, Warlop, and Wedel 2002; Pieters and Wedel 2004; Rosbergen, Pieters, and Wedel 1997; Wedel and Pieters 2000, 2008; Wedel, Pieters, and Liechty 2008), search effectiveness (Van der Lans, Pieters, and Wedel 2008b), and brand display on supermarket shelves (Chandon et al. 2009).

Other studies have used eye tracking in preference measurement settings. Toubia et al. (2012) use eye tracking in a purely descriptive manner to measure the impact of "gamifying" a preference measurement task on the amount of attention paid by consumers. Musalem, Meißner, and Huber (2013) employ eye tracking to explore how consumers' preferences for each level of an attribute relate to the amount of attention paid to that attribute level and to alternatives that contain it. Shi, Wedel, and Pieters (2013) use eye-tracking data to study and model how consumers switch back and forth between attribute-based and alternative-based strategies when acquiring information about products described in a matrix format. The research closest to ours is probably that of Stüttgen, Boatwright, and Monroe (2012): they develop and estimate a model of search and choice in which consumers are assumed to use a "satisficing" rule; that is, they evaluate alternatives one after another and choose the first alternative they deemed to be satisfactory. The authors further assume that consumers use a conjunctive rule to decide whether an alternative is satisfactory (i.e., all attributes of the alternative must be acceptable).

The standard approach for modeling eye-tracking data among these articles is either to treat eye fixations as exogenous (e.g., Musalem, Meißner, and Huber 2013) or to endogenize eye fixations using hidden Markov models (Liechty, Pieters, and Wedel 2003; Shi, Wedel, and Pieters 2013; Stüttgen, Boatwright, and Monroe 2012; Van der Lans, Pieters, and Wedel 2008a, b). The states in hidden Markov models of eye movements typically capture various information acquisition strategies or modes of search. For example, Stüttgen, Boatwright, and Monroe (2012) follow Liechty, Pieters, and Wedel (2003) and assume that consumers move back and forth between a "local search" state and a "global search" state, which involve eye movements in the periphery of the current eye position (local) and in different areas (global). Their model captures how the transition probabilities between these states are influenced by the consumer's ongoing evaluations of the various alternatives (i.e., which alternatives have already been classified as satisfactory or unsatisfactory on the basis of the information processed up to that point).

Our model takes a different approach. Compared with extant models based on hidden Markov processes, we allow consumers to be forward looking in how they acquire information. We model information acquisition as the result of forward-looking utility maximization, in which the utility a consumer derives comes not only from the chosen product but also from the information acquisition process itself. Another key feature of our model is that we allow for imperfect memory encoding; in other words, a consumer may need multiple fixations in a region of interest before remembering the information it contains.

MODEL

In this section, we develop a dynamic discrete choice model in which the information processed by consumers is endogenized and modeled as the result of forward-looking utility maximization, in which the consumer derives (positive or negative) utility both from his or her final choice and from the search process itself. The model is designed to be calibrated using a combination of eye-tracking and choice data.

Specification

For ease of exposition, we focus on one consumer when describing our model. We index choice questions by k, and each choice question consists of selecting one of J alternatives that are described by I attributes. For ease of presentation, we assume without loss of generality that all attributes have the same number of levels, L. The choice-relevant information is presented in a matrix such as that shown in Figure 1, with one column per alternative and one row per attribute.

For simplicity, we assume that the choice questions come from a random experimental design. In this case, attributes vary independently across alternatives, and there is no need to model inferences consumers may make across attributes and alternatives. However, our approach could easily be extended to nonrandom experimental designs.

Each time period in our model is a search occasion, t, in which the consumer chooses between acquiring some choice-relevant information (by visiting a cell in the matrix that contains the level of one attribute for one choice alternative) and ending the search and choosing one of the alternatives on the basis of the information acquired up to that point. In the latter case, the consumer moves to the next choice question.

As mentioned previously, our model is inspired by the DC model proposed and validated by Gabaix et al. (2006). According to that model, on each search occasion t, the participant chooses as if the search occasion were the last one in that question. In other words, if the consumer decides to acquire some information, he or she does so as if (s)he were going to make a choice immediately after acquiring this new piece of information.

We develop a likelihood-based implementation of the DC model that allows for heterogeneity in preferences. Like any dynamic model, our implementation specifies an action space, a set of state variables, a utility function, and state-transition probabilities. Next, we define each of these components and the resulting likelihood function.

Actions. We denote the current position of the eyes in the I × J matrix that contains the choice-relevant information by p = (i, j). On each search occasion, the consumer may move his or her eyes to a different location (i′, j′) in the matrix or end the information acquisition process and choose one of the alternatives (j′), thereby moving to the next choice question.[2]

States. Although the attribute levels for each alternative in a choice question are known to the researcher, they are unknown to the consumer at the beginning of the question. Take Figure 1 as an example. Consumers learn the level of each attribute in each alternative when they move their eyes to the relevant cell in the matrix. In our data, we found that consumers revisited 62.22% of the cells they visited at least once. Therefore, it would be unreasonable to assume that consumers learn the level of attribute i for alternative j with certainty after only one fixation in cell (i, j). Instead, we assume an imperfect memory encoding process in which consumers form a set of beliefs about the true value of each cell. These beliefs are updated after each fixation, and they converge to the truth as the number of fixations increases.

Our two observed state variables are p, which captures the current eye position, and a set of numbers $\{n_{i,j}\}$ in which $n_{i,j}$ is the number of times cell (i, j) was visited (i.e., number of fixations in the cell that contains information on attribute i for alternative j). We follow Wedel and Pieters (2000) and assume that consumers extract a chunk of information with each fixation on a cell. We denote as η the amount of information extracted per fixation. Again following Wedel and Pieters (2000), we further assume that the total amount of information stored by the consumer related to cell (i, j) is the sum of the activation levels of all memory traces: $\eta \times n_{i,j}$. Suppose the true level in cell (i, j) is l_0. After $n_{i,j}$ fixations in that cell, the total amount of information in support of l_0 being the true level is $\eta \times n_{i,j}$. The total amount of information in support of any other level being the true level is 0. If we assume some error in memory retrieval ($\delta_{i,j,l}$) (Wedel and Pieters 2000), the probability that the consumer believes l_0 is the true level is $\text{Prob}(\eta n_{i,j} + \delta_{i,j,l_0} > \delta_{i,j,l}, \forall\, l \neq l_0)$. If we assume that $\delta_{i,j,l}$ follows an i.i.d. double exponential distribution, the probability weight associated with each level l, $w_{i,j,l}$, becomes

$$(1)\quad w_{i,j,l}(\eta, n_{i,j}) = \begin{cases} \dfrac{\exp(\eta n_{i,j})}{L - 1 + \exp(\eta n_{i,j})} & \text{if } l \text{ is the true level} \\ \dfrac{1}{L - 1 + \exp(\eta n_{i,j})} & \text{if } l \text{ is not the true level} \end{cases}.$$

[2]We only consider fixations within the regions of interest that contain choice-relevant information. In the first search occasion in each question, the number of possible cells to move to is I × J instead of I × J − 1 (there is no "current" position). We collapse consecutive fixations within the same cell as one fixation because they are likely to be caused by participants randomly moving their eyes in a very small range, as a result of blinking (nonconsecutive fixations in the same cell are recorded as distinct fixations).

We denote the 1 × L array of probability weights corresponding to all possible levels in cell (i, j) as $w_{i,j}(\eta, n_{i,j})$, which equals $[w_{i,j,1}(\eta, n_{i,j}), \ldots, w_{i,j,L}(\eta, n_{i,j})]$. Before the first visit to a cell, the consumer starts with a uniform belief, $w_{i,j}(\eta, 0) = [1/L, \ldots, 1/L]$ that reflects the random experimental design (nonrandom designs would potentially give rise to different initial beliefs). The set of weights corresponding to a cell is updated after each visit to the cell and converges to a vector that has a weight of 1 on the true level and 0 on all the other levels. Appendix A illustrates this process using a simple example.

We further assume the existence of unobserved state variables in the form of idiosyncratic shocks ε(a) that capture information unobservable to the econometrician. This addition allows us to write a likelihood function for our model following a standard distributional assumption (see Rust 1987).

Utility function. The utility derived by the consumer at each search occasion is a function of the current state and the consumer's action. We make a distinction between product-related utility derived by the consumer (i.e., the utility that comes from the alternative j chosen by the consumer) and search-related utility (i.e., the utility that comes from the search process, which may be positive or negative). The consumer derives product-related utility only upon ending the search.

We first describe product-related utility. For ease of exposition, we do not include a subscript for the consumer in our equations, although all of the parameters are estimated at the individual level. We assume effects coding; that is, the partworth for the last level of an attribute is equal to minus the sum of the partworths for the other levels. Let β_i be the (L − 1) × 1 vector containing a consumer's partworths for attribute i under effects coding. The L × 1 vector containing all partworths for attribute i is $\tilde{\beta}_i = I_i^0 \beta_i$, where I_i^0 is an L × (L − 1) coding matrix:

$$(2)\quad I_i^0 = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ 0 & 0 & \ldots & 1 \\ 1 & \ldots & \ldots & -1 \end{bmatrix}.$$

With perfect memory encoding, the consumer would know the true level contained in cell (i, j) after one fixation, and the partworth corresponding to that cell would simply be the appropriate element of $\tilde{\beta}_i$. However, with imperfect memory encoding, the consumer assigns a set of probability weights, $w_{i,j}(\eta, n_{i,j})$, to each possible level in cell (i, j), and the expected value of the partworth corresponding to that cell is a weighted average of the partworths for attribute i: $w_{i,j}(\eta, n_{i,j})\tilde{\beta}_i$. This expression converges to the appropriate element of $\tilde{\beta}_i$ as the number of fixations on the cell increases. The product-related utility is specified as

$$(3)\quad u_{\text{product}}(a|\{n\}, \beta) = \begin{cases} 0 & \text{if } a = \text{move to } (i', j') \\ \sum_i w_{i,j'}(\eta, n_{i,j'})\tilde{\beta}_i & \text{if } a = \text{choose } j' \end{cases}.$$

Appendix A illustrates computation of product-related utility using a simple example.

In a case in which the consumer decides not to choose any alternative and continues searching instead, he or she derives search-related utility. Given the finding that the amount of information processed by participants tends to decrease as they progress through the questionnaire (Stüttgen, Boatwright, and Monroe 2012; Toubia et al. 2012), we allow for fatigue effects by modeling search-related utility as a function of the question number, k.

Shi, Wedel, and Pieters (2013) document that a large proportion of consecutive fixations are to contiguous cells when information about products is displayed using matrices, as in our case. This is consistent with eye movements between cells being more cognitively costly when the cells are more distant. Moreover, Shi, Wedel, and Pieters identify an asymmetry in consumers' propensity to make horizontal versus vertical eye movements. To capture these physiological factors, we also allow search-related utility to be a function of the distance between the current location of the eyes, p, and the next cell visited and allow for different weights on horizontal and vertical movements. In particular, we model search-related utility as follows:[3]

$$(4)\quad u_{search}(a|p,k,\theta)=\begin{cases}\theta_0+\theta_1 k+\theta_2 d(a,p,\theta_3) & \text{if } a=\text{move to } (i',j')\\ 0 & \text{if } a=\text{choose } j'\end{cases},$$

where $d(a, p, \theta_3)$ is a weighted Euclidean distance between the current cell (i, j) and the next cell (i′, j′) defined as $\sqrt{(i-i')^2+\theta_3(j-j')^2}$. The parameter θ_3 captures asymmetries between vertical and horizontal eye movements (this parameter is constrained to be nonnegative). Note that we do not restrict the signs of the parameters θ_0, θ_1, and θ_2.

Transition probabilities. The state variables capture eye position and the number of fixations in each cell. The transitions between states are deterministic from the perspective of the researcher, given the customer's actions. However, from the perspective of the consumer, there is some uncertainty regarding the true value of each cell, so the transitions between states are probabilistic. Suppose that the consumer's action, a, is to visit cell (i′, j′). From the perspective of the consumer, for each possible level l there is a probability (given by $w_{i',j',l}[\eta, n_{i',j'}]$) that this level will be found in cell (i′, j′). Therefore, the value function is based on a set of transition probabilities given by the current set of probability weights, $w_{i',j'}(\eta, n_{i',j'})$, which are a function of the current state variable, $n_{i',j'}$. These weights represent the probability of observing each level in each cell on the basis of the consumer's current beliefs. (Note that we assume that the consumer knows which cell he or she is visiting; the only uncertainty is related to the level contained in the cell.)

Suppose that l_0 is the true level in cell (i′, j′). To specify the Bellman equation, we need to model all possible state transitions from the perspective of the consumer. In particular, we need to model what would happen when the consumer sees a level in cell (i′, j′) that is different from l_0. Although this never happens, it must be addressed because the weights $[w_{i',j',l}(\eta, n_{i',j'})]$ are positive for all levels, not just the true level. If level $l \neq l_0$ were to be found in cell (i′, j′), the amount of information in support of that level being the true level would increase from 0 to η. As a result, the probability weight associated with the level would be updated to $\exp(\eta)/[L-2+\exp(\eta n_{i',j'})+\exp(\eta)]$. The weight associated with the true level would be updated to $\exp(\eta n_{i',j'})/[L-2+\exp(\eta n_{i',j'})+\exp(\eta)]$, and the weights associated with the other levels would be updated to $1/[L-2+\exp(\eta n_{i',j'})+\exp(\eta)]$. With a slight abuse of notation, we denote the product-related utility based on the new beliefs that would be formed if level l were found in cell (i′, j′) as $u_{product}(a'|\{n'_{i',j',l}\}, \beta)$.

After the fixation to cell (i′, j′), the state variable corresponding to that cell, $n_{i',j'}$, is incremented by 1, and the set of probability weights corresponding to that cell is updated to $w_{i',j'}(\eta, n_{i',j'}+1)$ on the basis of Equation 1. There is no need to specify transition probabilities for the other state variables (p[current fixation position] and ε[a]) because the former evolves deterministically and the latter is assumed to satisfy the conditional independence assumption (Rust 1987).

Likelihood function. The DC model assumes that consumers act on each search occasion as if this search occasion were their last opportunity to acquire new information. Mathematically, this implies that consumers behave on each search occasion as if they are solving the following optimization problem:

$$(5)\quad \max\left\{\max_{a=\{j'\}}\left[u_{product}(a|\{n\},\beta)+\varepsilon(a)\right],\ \max_{a=\{i',j'\}}\left(u_{search}(a|p,k,\theta)+\varepsilon(a)+\sum_{l}w_{i',j',l}(\eta,n_{i',j'})\times\max_{a'=\{j\}}\left[u_{product}(a'|\{n'_{i',j',l}\},\beta)+\varepsilon(a')\right]\right)\right\}$$

The first term, $\max_{a=\{j'\}}\{u_{product}(a|\{n\}, \beta)+\varepsilon(a)\}$, is the maximum utility the consumer can derive by ending the search and choosing one of the alternatives given the current state variables {n} and the consumer's partworths β. The second term is the maximum utility the consumer can derive by continuing the search, where $u_{search}(a|p, k, \theta)+\varepsilon(a)$ is the search-related utility and $w_{i',j',l}(\eta, n_{i',j'})$ captures the state-transition probabilities (from the consumer's perspective) and $\max_{a'=\{j\}}[u_{product}(a'|\{n'_{i',j',l}\}, \beta)+\varepsilon(a')]$ is the maximum utility derived from choosing one of the alternatives in the next period given that level l is found in cell (i′, j′).

Assuming that the idiosyncratic shocks, ε, satisfy the conditional independence assumption and follow a double-exponential distribution gives rise to the following likelihood function in which $\Theta=\{\beta, \theta, \eta\}$:

$$(6)\quad P(a|\{n\},p,\Theta)=\frac{\exp\left[V_a(\{n\},p|\Theta)\right]}{\sum_{a'}\exp\left[V_{a'}(\{n\},p|\Theta)\right]},$$

where

[3]We also tested a version of the model with only the parameter θ_0 in the search-related utility function. The deviance information criteria (DICs) favored the more complex version, and out-of-sample performance was not greatly affected. Details are available from the authors.

$$(7)\qquad V_a(\{n\}, p|\Theta) = \begin{cases} u_{search}(a|p,k,\theta) + \sum_l w_{i',j',l}(\eta, n_{i',j'}) \\ \times \log \sum_{a'=\{j\}} \exp\left[u_{product}\left(a'\middle|\left\{n'_{i',j',l}\right\},\beta\right)\right] & \text{if } a = \text{move to } (i',j') \\ u_{product}(a|\{n\},\beta) & \text{if } a = \text{choose } j' \end{cases}$$

Identification and Estimation

The parameters to be estimated in our proposed dynamic discrete choice model are $\Theta = \{\beta, \theta, \eta\}$. As with a standard CBC analysis, the partworths, β, are identified at the individual level through the choices consumers make between various alternatives. The parameters θ_0, θ_1, θ_2, and θ_3 capture search-related utility. Given the value of the partworths, the intercept θ_0 is identified because we observe consumers who choose either to continue the search or to stop the search and select one of the alternatives. The parameter θ_1 captures the effects of fatigue (through the question number) and is identified because we observe multiple questions per consumer. The parameters θ_2 and θ_3 capture the effect of distance on search utility and are identified because the information in each cell varies randomly across questions. We estimate both β and θ at the individual level. Finally, η is a parameter that captures the amount of information extracted per fixation (i.e., it may be interpreted as capturing the speed with which consumers learn the content of a cell). This parameter is identified primarily through the common occurrence of revisits to cells that the same consumer previously visited in the same question. Although this parameter is parametrically identified in theory, we find that it is only weakly identified in practice. Intuitively, it is difficult to disentangle at the level of each consumer the extent to which revisits are driven by slow learning (low value of η) versus strong preferences (i.e., cells that are revisited frequently tend to correspond to more important attributes). Our experience, based on real and simulated data, suggests that this problem is not specific to any particular data set. However, we have found that this parameter is adequately identified at the aggregate level. Therefore, we estimate it at the aggregate level using a grid search. In particular, we fix the parameter η and estimate all the other parameters given that value of η for multiple values of η. We keep the value of η that gives rise to the lowest deviance information criteria (DICs).[4]

We estimate our model using a hierarchical Bayes method (Atchadé and Rosenthal 2005). The first-stage prior for $\{\theta_n, \beta_n\}$ (where n indexes consumers) is normal with $\{\theta_n, \beta_n\} \sim N(\mu_0, \Lambda)$. The second-stage priors are $\mu_0 \sim N(0, 1{,}000 \times I)$ and $\Lambda^{-1} \sim \text{Wishart}(I, 23 + 3)$, where 23 is the number of heterogeneous parameters in the model. A total of 150,000 Markov chain Monte Carlo (MCMC) iterations are performed using the first 100,000 as burn-in. We apply a grid search method for the learning parameter η; we estimate the model with $\eta = 0$–5 with a step of 1 and select the best-fitting model on the basis of the DICs. The Web Appendix provides details of our estimation procedure.

We confirmed the identification of our model and tested our estimation approach using a simulation study. Appendix B provides the details of our simulation study. We generated a synthetic data set using a set of parameters inspired by the estimates from our study reported in the "Estimation Results" subsection. We found that the parameters were recovered adequately.

Comparison with Gabaix et al. (2006)

We used Gabaix et al.'s (2006) DC model as a basic framework for our model, but our model differs significantly in several important ways. First, Gabaix et al.'s model was applied to a context in which each cell contained a monetary amount, and the payoff from the chosen alternative was the sum of the monetary values of its cells. Product-related utility in Gabaix et al.'s model was simply the amount of money earned in the game. We apply our model to a context in which each cell contains an attribute level and product-related utility is parameterized by a set of partworths. Second, Gabaix et al. assumed that the value in each cell was drawn from a continuous normal distribution, whereas in our case, the values are drawn from a discrete uniform distribution. As a result, Gabaix et al. were able to derive a closed-form expression for the expected benefit from each possible action (see Equation 3 in Gabaix et al. [2006]), whereas we use the general Bellman equation. Third, Gabaix et al. assumed that search-related utility is constant (the opportunity cost of time), whereas we allow search-related utility to be affected by fatigue and proximity effects (i.e., consumers may search less over time and may be more likely to move their eyes to nearby cells). Fourth, Gabaix et al. assumed perfect memory encoding (i.e., a consumer learns the content of a cell perfectly after one visit), whereas we allow for imperfect memory encoding. Fifth, Gabaix et al. calibrated the one parameter in their model (opportunity cost of time) by fitting moments of the data (average amount of search in the game); we develop a likelihood-based hierarchical Bayesian framework. Sixth, Gabaix et al. calibrated their model at the aggregate level; we allow for heterogeneity across consumers.

DATA

Setup

We collected CBC data in the context of Dell laptop computers and used six attributes (I = 6) with four levels each (L = 4): processor speed (1.6 GHz, 1.9 GHz, 2.7 GHz, and 3.2 GHz), screen size (26 cm, 35.6 cm, 40 cm, and 43 cm), hard drive capacity (160 GB, 320 GB, 500 GB, and 750 GB), Dell support subscription (1 year, 2 years, 3 years, and 4 years), McAfee antivirus subscription (30 days, 1 year, 2 years, and 3 years), and price (350€, 500€, 650€, and 800€). In the main task, each participant answered 20 choice questions, each offering four alternatives (J = 4). The questions were generated randomly (once for all participants; i.e., all participants saw the same set of questions). Before answering the 20 questions, participants completed a training question designed to familiarize them with the interface. Figure 1 provides a screenshot of a choice question.

[4]The DIC is defined as $-4E_\Theta[\log P(\{a\}|\{n\}, \{p\}, \Theta)] + 2 \log P(\{a\}|\{n\}, \{p\}, \hat{\Theta})$, where $P(\{a\}|\{n\}, \{p\}, \Theta)$ is the likelihood function and $\hat{\Theta} = \arg\max_\Theta[P(\{a\}|\{n\}, \{p\}, \Theta)]$ (Celeux et al. 2006).

In addition to the main task, participants completed an external validity task. We used a typical setting in which the external validity task consisted of a choice task with eight alternatives that were chosen randomly (once for all participants; i.e., all participants saw the same set of alternatives) subject to the constraint that each level of each attribute would be present in at least one of the alternatives. This task also was preceded by a training question to familiarize participants with the interface.

We randomized the position of the external validity task relative to the main task so that half the participants completed the external validity task first and the other half completed the main task first. This difference was our only between-subjects variation.

Our study followed an incentive-alignment scheme typical of CBC studies (Ding 2007; Ding, Grewal, and Liechty 2005). One participant was selected randomly as a winner and received 800€, which was used to automatically purchase a laptop based on his or her answers to the survey. The winner received the alternative chosen in the external validity task with probability 50% and the alternative chosen in each question in the main task with probability 2.5%. The winner received the preferred laptop along with the difference between 800€ and the price of the laptop.

Our participants were recruited at a large European university. They all participated in the survey in the university's behavioral lab using the online platform developed by the authors.

Participants completed the survey while being monitored by a free-standing nonintrusive Tobii 2150 eye tracker that sampled infrared corneal reflections at 50Hz with a .35-degree spatial resolution and an accuracy of .5 degrees. The stimuli were presented on a 21-inch LCD monitor with a display resolution of 1600 × 1200 pixels. The position of the left eye and right eye were recorded separately (Van der Lans, Wedel, and Pieters 2011). Fixations and saccades were differentiated using Van der Lans, Wedel, and Pieters's (2011) velocity-based algorithm. We defined the region of interest for each piece of information as the area within the boundary of the cell that contains the information (see Figure 1).

Descriptive Statistics

We collected complete eye-tracking data for 70 participants,[5] of whom 33 completed the external validity task before the main task and 37 completed it after the main task. We next provide a descriptive analysis of our eye-tracking data for the 20 questions in the main task. The average proportion of cells visited at least once (with at least one fixation) across all questions and participants was 69.65%. The proportion differs slightly with the order in which the main task and the external validity task were completed: 67.79% for the main task first and 71.74% for the external validity task first ($p = .13$). We accommodated this difference by adding a parameter to our search-utility specification (see the "Proposed Model" subsection). Figure 2 plots the average proportion of information visited in each choice question. The downward trend in this graph confirms the need to control for question position in our model and is consistent with previous findings (Stüttgen, Boatwright, and Monroe 2012; Toubia et al. 2012). Figure 3 shows the distribution of the proportion of information visited across all choice questions and participants. Figure 4 shows the distribution of the num-

Figure 2
AVERAGE PROPORTION OF INFORMATION VISITED PER CHOICE QUESTION VERSUS QUESTION NUMBER

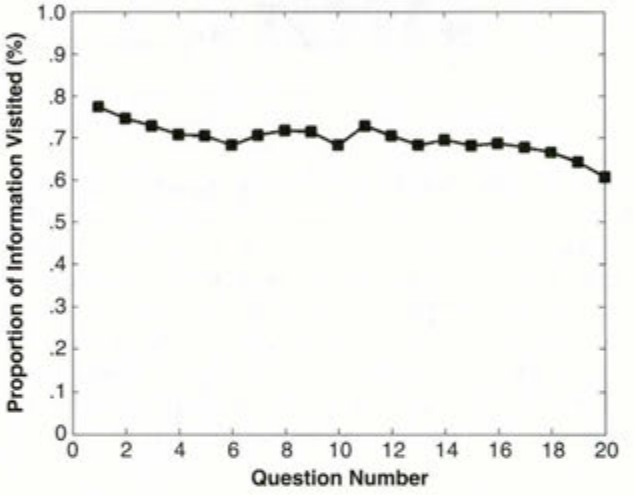
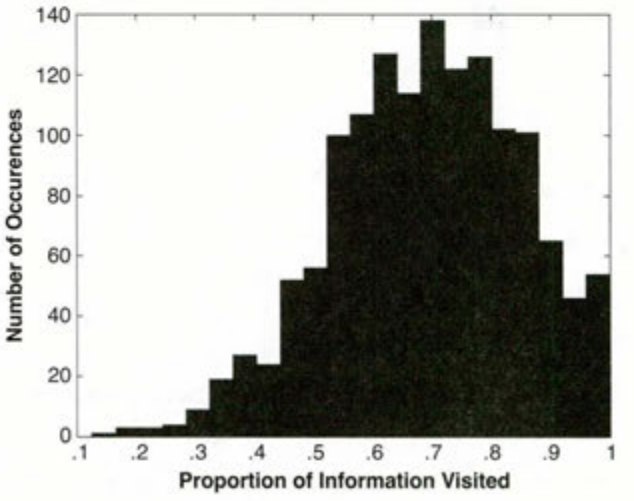
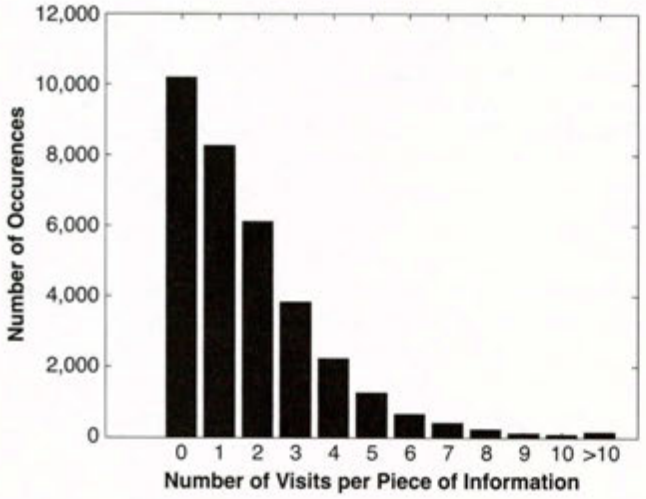


Figure 3
DISTRIBUTION OF THE PROPORTION OF INFORMATION VISITED PER CHOICE QUESTION (ACROSS ALL RESPONDENTS AND CHOICE QUESTIONS)



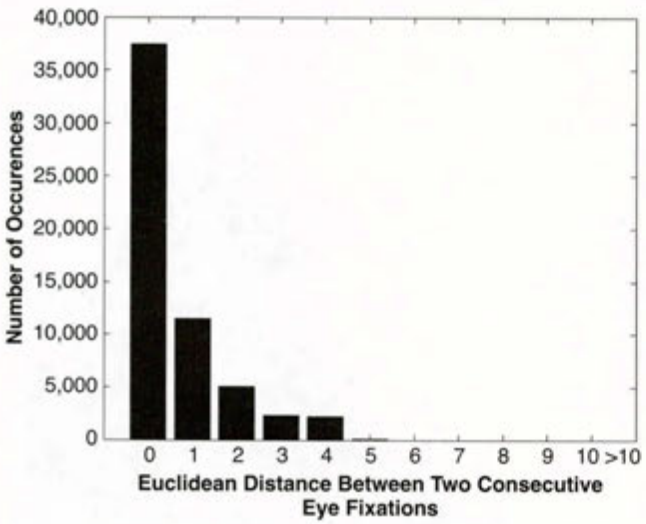
[5]We recruited 120 respondents to participate in our study. From the raw eye-fixation data, we identified time stamps without any affiliation of eye-fixation position as missing data. The respondents with missing data were confirmed through a video of their eye movements mapped onto the choice experiment interface. This led us to exclude 50 of the 120 respondents. Our large proportion of incomplete respondents was due to the eye tracker being near the end of its life (it was decommissioned a few weeks after we finished our study). However, we have no reason to believe that data were missing nonrandomly or that data were recorded incorrectly for our complete respondents. The large proportion of missing data reduced our statistical power but should not change our results.

Figure 4
DISTRIBUTION OF THE NUMBER OF VISITS PER PIECE OF INFORMATION (ACROSS ALL PIECES OF INFORMATION, RESPONDENTS, AND CHOICE QUESTIONS)



ber of visits per piece of information (each "piece of information" consists of the level of one attribute for one alternative) for all pieces of information, choice questions, and participants. This chart shows that information that is processed is likely to be visited multiple times by the same consumer in the same question, which confirms the need to model memory encoding as imperfect; it would not be reasonable to assume that visiting a cell once is enough for a consumer to completely memorize its content. Figure 5 shows the distribution of the distances between two consecutive eye fixations across all choice questions and participants.[6]

Figure 5
DISTRIBUTION OF THE EUCLIDEAN DISTANCE BETWEEN SUCCESSIVE EYE FIXATIONS



Notes: If the respondent moves his or her eyes from cell (i, j) to cell (i', j'), the distance is defined as $\sqrt{(i-i')^2+(j-j')^2}$.

We observe that consumers are much more likely to move their eyes to an adjacent (distance = 1) cell in the 6 × 4 matrix containing all choice-relevant information than they are to move a more distant cell. This is consistent with previous studies (Shi, Wedel, and Pieters 2013; Stüttgen, Boatwright, and Monroe 2012) and confirms the need to model search-related utility as a function of the distance between cells.

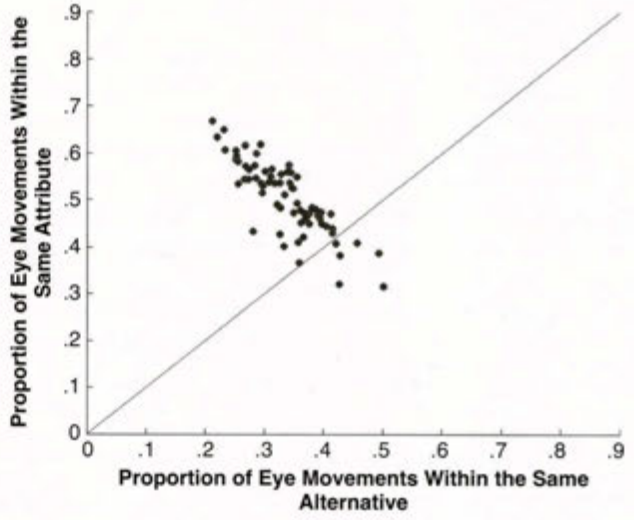
Table 1 shows the proportion (across all questions and participants) of eye movements that were to a different choice alternative within the same attribute, to a different attribute within the same alternative, and to a different attribute in a different alternative. Although most movements were either within the same alternative or within the same attribute, there is no evidence that either alternative-based processing or attribute-based processing dominates. To explore the possibility that each type of processing dominates for subsets of consumers, we present a scatterplot of the proportion of within-attribute and within-alternative eye movements at the participant level in Figure 6 (each dot rep-

[6]If the respondent moves his or her eyes between cell (i, j) and (i', j'), the distance is defined as $\sqrt{(i-i')^2+(j-j')^2}$.

Table 1
OVERALL PROPORTION OF DIFFERENT TYPES OF EYE MOVEMENTS

Type of Search	*Proportion*
Different alternative within the same attribute	.51
Different attribute within the same alternative	.34
Different attribute in a different alternative	.16

Figure 6
SCATTERPLOT OF THE PROPORTION OF EYE MOVEMENTS TO A DIFFERENT ALTERNATIVE WITHIN THE SAME ATTRIBUTE VERSUS A DIFFERENT ATTRIBUTE WITHIN THE SAME ALTERNATIVE



Notes: Each dot corresponds to one respondent.

resents one participant). We find that most participants use a hybrid of attribute-based and alternative-based searches, although attribute-based searches were more prevalent, on average. To further investigate the existence of attribute-based and alternative-based searches, we report the distribution of the number of attributes visited per alternative (across all alternatives, respondents, and choice questions) in Figure 7 and the number of alternatives visited per attribute (across all attributes, respondents, and choice questions) in Figure 8. Attribute-based search would lead to some attributes not being visited at all, and alternative-based search would lead to some alternatives not being visited at all. We find that an alternative (attribute) is completely ignored only 1.13% (4.33%) of the time, which further suggests that no evidence exists for purely attribute-based or alternative-based searches and confirms the need for a model to be flexible enough to allow for any type of eye movements.

Figure 7
DISTRIBUTION OF THE NUMBER OF ATTRIBUTES VISITED PER ALTERNATIVE (ACROSS ALL ALTERNATIVES, RESPONDENTS, AND CHOICE QUESTIONS)

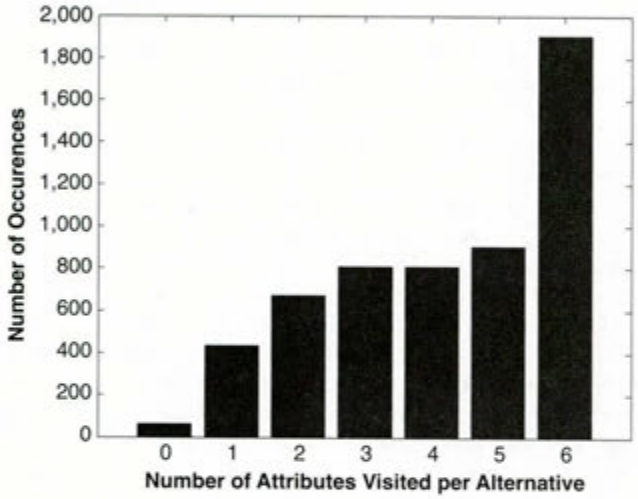
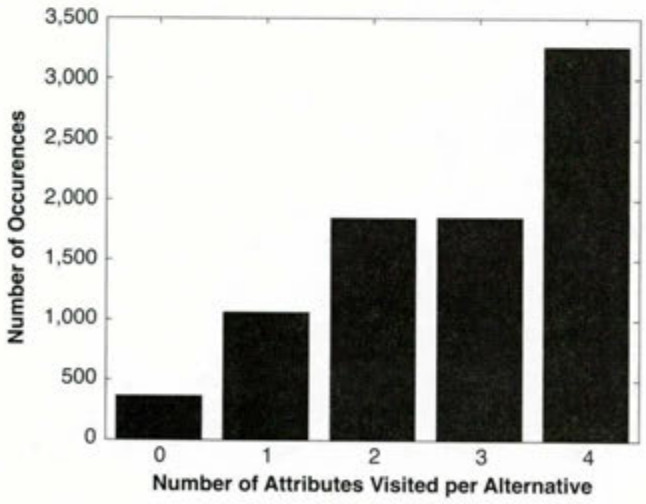


Figure 8
DISTRIBUTION OF THE NUMBER OF ALTERNATIVES VISITED PER ATTRIBUTE (ACROSS ALL ATTRIBUTES, RESPONDENTS, AND CHOICE QUESTIONS)



Finally, we explore how consumers divide their attention among the various attributes across choice questions. Figure 9 plots the distribution of the share of fixations for each attribute (i.e., the proportion of fixations in cells containing information for that attribute), across choice questions. We observe that the shares do not vary much across questions but that there are signification variations across attributes. On average, processor speed attracts the most fixations (27%), followed by screen size (21%), price (18%), hard drive capacity (17%), Dell support subscription (9%), and McAfee antivirus subscription (8%).

ESTIMATION RESULTS

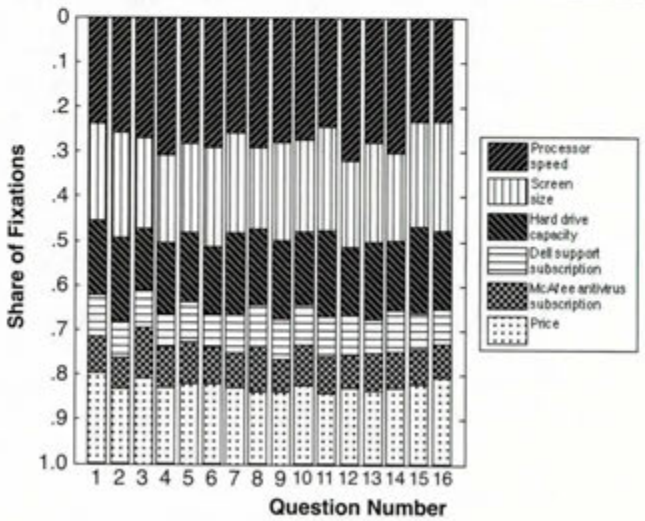
We use the last 4 questions of the main task as holdouts. We vary the number of questions used for estimation between the first 8 and 16 to assess the benefits of the proposed model when the number of choice questions is increased.

Proposed Model

We estimate the proposed model described previously, with one small adjustment to capture the change in position of the external validity task and its effect on the propensity to search. Thus, we add one term to the search-related utility equation:

$$(8)\qquad u_{search}(a|p,k,\theta) = \begin{cases} \theta_0 + \theta_{11} + \theta_{12}\mathbb{1}(\text{ext val}) + \theta_2 d(a,p,\theta_3) & \text{if } a = \text{move to } (i',j') \\ 0 & \text{if } a = \text{choose } j' \end{cases}.$$

Figure 9
AVERAGE SHARES OF FIXATIONS PER ATTRIBUTE (ACROSS RESPONDENTS) VERSUS QUESTION NUMBER



where $\mathbb{1}$(ext val) is an indicator function equal to 1 if the participant completed the external validity task before the main task. The additional parameter θ_{12} captures the effect of completing the external validity task before the main task on the propensity to search for information in the main task. Recall that we omit consumer subscripts for ease of exposition, but the partworths and search-related utility parameters are all estimated at the individual level. In addition, we constrain the partworths for price to be monotonic using rejection sampling (Allenby, Arora, and Ginter 1995).

Benchmark Models

We use several benchmark models to test various components of the proposed model. All benchmarks are estimated using the same Bayesian approach and the same prior specifications as the proposed model.

Our first set of benchmarks does not model search and focuses only on choice. We refer to this set as the "choice-only" benchmarks. The first is a standard multinomial logit choice model that assumes that participants have full knowledge of the alternatives in each choice question (i.e., the information in all of the 24 cells in Figure 1 is assumed to be known). This benchmark assumes that the utility of choosing alternative j is simply the sum of the partworths associated with each attribute, plus a random error term that is i.i.d. extreme value. The second benchmark is a multinomial logit model that takes into account information on the specific cells visited (cells with at least one fixation) by each participant in each question but assumes that search is exogenous. This benchmark assumes that participants only use the information contained in the cells that they visited when they evaluate the alternatives. That is, it assumes that the utility of choosing alternative j is the sum of the partworths associated with the cells that were visited at least once for this alternative, plus a random error term that is i.i.d. extreme value. These two benchmarks only model choice (i.e., their likelihood functions only capture how well they fit the choice data). Therefore, we cannot compare them with the proposed model using measures of in-sample fit, so we use out-of-sample predictions instead.

Our second set of benchmarks models both the choices made by consumers and their eye movements and therefore may be compared with our proposed model using in-sample fit statistics (DICs). We refer to this set as "search+choice" benchmarks. In each search+choice benchmark, the same imperfect memory encoding process is assumed as in the proposed model (Equation 1), and the same specification is used for product-related utility (Equation 3) and search-related utility (Equation 4). The only difference is in the specification of the forward-looking term in the value function (Equation 7). The first benchmark in this set (labeled "future product-related utility unanticipated") assumes that consumers only take search-related utility into consideration when deciding whether and how to search for information and that they ignore future product-related utility. This benchmark assumes that the value function from Equation 7 takes the following form:

$$(9)\quad V_a(\{n\},p|\Theta)=\begin{cases} u_{search}(a|p,k,\theta) & \text{if } a=\text{move to } (i',j') \\ u_{product}(a|\{n\},\beta) & \text{if } a=\text{choose } j' \end{cases}.$$

Our second benchmark in this set explores the possibility that although consumers may take future utility into account when deciding whether and how to search for information, they may not take into account how the results of the search will affect their future beliefs, which will influence their future expected utility. This benchmark assumes that consumers behave on each search occasion as if they will not update their beliefs after the search. We label this benchmark as "future belief updating unanticipated." Note that this benchmark assumes that participants ignore future updating of beliefs at the time of the decision, but beliefs are updated after the new information is acquired. The value function takes the following form:

$$(10)\quad V_a(\{n\},p|\Theta)=\begin{cases} u_{search}(a|p,k,\theta) \\ \quad +\log\sum_{a'=\{j\}}\exp\left[u_{product}(a'|\{n\},\beta)\right] & \text{if } a=\text{move to } (i',j') \\ u_{product}(a|\{n\},\beta) & \text{if } a=\text{choose } j' \end{cases}.$$

The term $\log\Sigma_{a'=\{j\}}\exp[u_{product}(a'|\{n\},\beta)]$ in Equation 10 replaces $\Sigma_l w_{i',j',l}(\eta,n_{i',j'})\log\Sigma_{a'=\{j\}}\exp[u_{product}(a'|\{n'_{i',j',l}\},\beta)]$ from Equation 7. That is, the anticipated product-related utility in the next period is based on the current state variables $\{n\}$ instead of being based (probabilistically) on the state variables in the next period, $\{n'_{i',j',l}\}$.

We estimate η, the learning parameter, for each benchmark separately using a grid search of the same set of values as in the proposed model. Table 2 reports the DICs for the proposed model and the second set of benchmarks when the number of questions used for calibration varies from 8 to 16. We find that the proposed model has a better fit than both benchmarks. Thus, it seems reasonable to assume that consumers take future product utility into account when deciding whether and how to search and that they anticipate how searching will affect their beliefs about the various alternatives.

Posterior Check

In addition to computing the DICs, we measure in-sample fit for the proposed model by evaluating how well it recovers some key statistics of the data (Gelman, Meng, and

Table 2
COMPARISON OF PROPOSED MODEL WITH SEARCH+CHOICE BENCHMARKS BASED ON DICs

Number of Questions Used for Calibration	*Proposed Model*	*Future Belief Updating Unanticipated*	*Future Product-Related Utility Unanticipated*
8	140,758.52	142,849.95	143,390.12
9	156,895.06	159,134.89	159,746.14
10	172,595.80	175,173.67	175,816.28
11	190,780.44	193,607.85	194,222.65
12	207,433.84	210,737.01	211,265.65
13	222,558.98	226,183.09	226,766.77
14	236,862.55	240,554.61	241,243.02
15	252,007.35	255,871.44	256,665.68
16	267,657.97	271,813.86	272,617.37

Stern 1996). At each iteration of the Gibbs sampler, we use the parameter estimates in that iteration to simulate the number of eye fixations for each respondent and for each choice question in the calibration set and repeat this analysis when the number of questions used for calibration varies from 8 to 16. Figure 10 shows the real average number of eye fixations across choice questions and respondents together with the 95% credible intervals of this statistic across iterations of the Gibbs sampler as the number of questions used for calibration varies. In all cases, the true values fall within the 95% credible interval.

Parameter Estimates

Table 3 shows the posterior means and 95% credible intervals of the first-stage prior parameters defined in the "Identification and Estimation" subsection—that is, the population mean of the partworths and search-utility parameters (μ_0) and the population variance of those parameters (diagonal elements of Λ). The estimates presented in Table 3 are based on the proposed model using the first 16 questions for calibration. (Recall that all parameters in Table 3 are estimated at the individual level; we report only the population means here.) All results are based on $\eta = 3$, the value suggested by the grid search for this parameter.[7] As previously mentioned, we used effects coding such that the partworths sum to 0 within each attribute.

The signs of θ_{11} and θ_{12} are consistent with the descriptive statistics reported previously that show that search decreases as the questionnaire progresses and that placing the external validity first slightly increases the amount of attention spent in the main task. The sign of θ_2 is consistent with the finding that consumers tend to move to cells that are close to the one they are currently visiting. The fact that θ_3 is less than 1 is consistent with Table 1, which shows that horizontal eye movements within the same attribute are more frequent than vertical ones within the same alternative. The positive sign of θ_0 suggests that search-related utility may be positive in certain situations.

Table 4 presents the posterior means of the importance of each attribute based on the proposed model and the set of choice-only benchmarks. The proposed model extracts more information about each attribute from each choice question. Therefore, we expect this model to give rise to more discrimination across attributes; for each consumer, there should be more variance in attribute importance across attributes. Moreover, attribute importances should be correlated with the number of fixations in each attribute, with consumers spending more attention on more important attributes. Table 4 reports posterior means and credible intervals of the averages (across consumers) of the variance (across attributes) of the partworth importances. That is, at each iteration of the MCMC, we compute the variance of the attribute importances for each consumer and calculate the average of this variance across consumers. As we predicted, the proposed model gives rise to much more variation across attributes than the benchmarks (the 95% credible

[7]The DIC for the proposed model when calibrating on the first 16 questions is 274,722.36; 268,874.46; 267,707.81; 267,657.96; 267,906.75; and 268,204.44 when varying η from 0 to 5 with a step of 1.

Figure 10
POSTERIOR CHECK OF THE AVERAGE NUMBER OF EYE FIXATIONS PER CHOICE QUESTION VERSUS NUMBER OF QUESTIONS USED FOR CALIBRATION

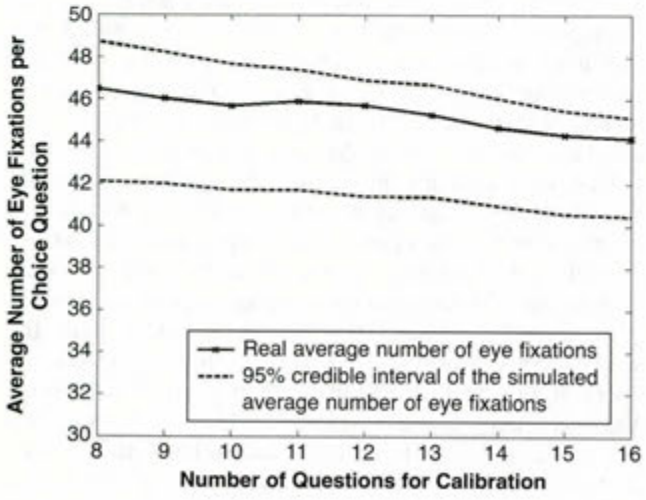


Table 3
POPULATION ESTIMATES FROM THE PROPOSED MODEL

	Posterior Population Mean	*95% Credible Interval*	*Posterior Population Variance*
Search-Related Parameters			
θ_0	2.33	[2.13, 2.51]	1.05
θ_{11}	−.01	[−.03, .00]	.36
θ_{12}	.28	[−.07, .61]	1.30
θ_2	−1.07	[−1.08, −1.05]	.45
θ_3	.70	[.66, .73]	.75
Processor Speed			
1.6 GHz	−8.80	[−9.11, −8.51]	17.03
1.9 GHz	−3.21	[−3.44, −2.97]	6.54
2.7 GHz	3.94	[3.69, 4.15]	5.07
Screen Size			
26 cm	−.31	[−.54, −.09]	17.06
35.6 cm	1.42	[1.16, 1.77]	4.36
40 cm	.31	[−.01, .51]	4.14
Hard Drive			
160 GB	−3.36	[−3.87, −2.99]	4.27
320 GB	−.62	[−1.12, −.28]	2.57
500 GB	1.20	[.98, 1.51]	2.68
Dell Support			
1 year	−1.51	[−1.75, −1.27]	1.93
2 years	.90	[.69, 1.11]	1.81
3 years	.16	[−.07, .41]	1.55
Antivirus			
30 days	−.83	[−1.03, −.57]	4.37
1 year	.05	[−.23, .39]	1.44
2 years	.58	[.31, .82]	2.13
Price			
350€	4.98	[4.68, 5.28]	14.18
500€	1.49	[1.34, 1.66]	1.97
650€	.04	[−.13, .17]	1.29

Notes: The first 16 questions are used for calibration. We used effects coding, so the partworth for the last level of each attribute is equal to minus the sum of the other three partworths.

Table 4
AVERAGE ATTRIBUTE IMPORTANCE AND VARIANCE OF ATTRIBUTE IMPORTANCE

		Choice-Only Benchmarks	
	Proposed Model	*Assume Consumers Fully Informed*	*Use Knowledge of Which Cells Visited*
Average attribute importance			
Processor speed	.335	.281	.232
Screen size	.145	.140	.140
Hard drive capacity	.140	.199	.189
Dell support subscription	.076	.091	.125
McAfee antivirus subscription	.081	.090	.108
Price	.222	.200	.206
Average variance of attribute importance	.010	.006	.003
95% credible interval	[.009, .011]	[.004, .007]	[.002, .003]

Notes: The first 16 questions are used for calibration.

intervals do not overlap with the benchmarks). We also observe a positive correlation between attention and importance: the rank order correlation between the average shares of fixations reported in Figure 9 and the average attribute importances reported in Table 4 is .89. This correlation may also be computed for each question for each respondent. The average is .67 (SD = .31) and the median is .76 across all choice questions and respondents. In other words, complementing choice data with search data and modeling the information acquisition process as the result of forward-looking utility maximization increases discrimination across attributes, and attributes that receive more attention tend to have larger estimated importances.

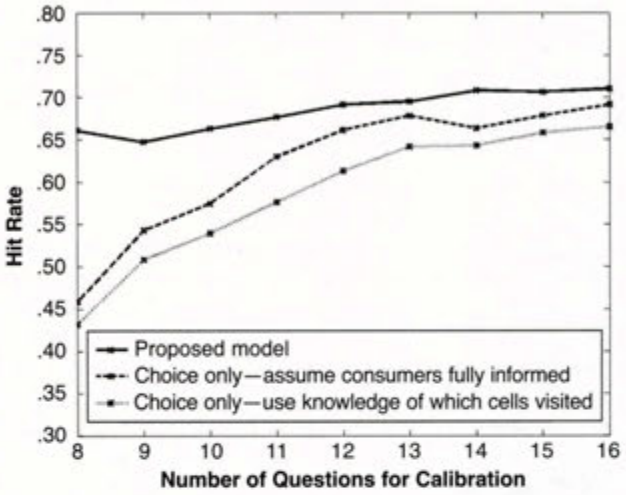
Out-of-Sample Predictions

We compare out-of-sample prediction performance between the proposed model and the two sets of benchmarks (search+choice and choice-only) using the hit rate on both the holdout questions (the last four questions in the main task) and the external validity task. For each consumer and out-of-sample question, we measure the hit rate by computing the estimated choice probability of the chosen alternative at each MCMC iteration and then compute the average across MCMC iterations.[8] We plot how the average performance of each model evolves as the number of questions used for calibration varies between 8 and 16. Figures 11–14 report the results.

We observe in Figure 11 that the hit rate on the holdout questions is systematically higher in the proposed model than in the choice-only benchmarks and that the difference becomes less pronounced as the number of questions

[8]We computed the utility of each alternative in each out-of-sample choice question by multiplying the characteristics of the alternatives by the consumer's partworths. This standard approach implicitly assumes that consumers take into account the full description of all the alternatives in the choice set. We also tried making out-of-sample predictions for the search+choice models based on counterfactual simulations. In particular, we could simulate the consumer's search process in the out-of-sample questions and estimate the resulting choice probabilities. Predictive performance was slightly worse using this approach. Details are available from the authors.

Figure 11
PROPOSED MODEL VERSUS CHOICE-ONLY BENCHMARKS: AVERAGE HOLDOUT HIT RATE VERSUS NUMBER OF QUESTIONS USED FOR CALIBRATION



increases. To compare the performance of the proposed model and each set of the benchmarks statistically, we run a set of ordinary least squares regressions with the logit of the hit rate as the dependent variable. Tables C1 and C2 in Appendix C present the results. We find that the main effects that correspond to the two choice-only benchmarks are significantly negative; the proposed model, on average, performs significantly better than either benchmark. This is consistent with the fact that the proposed model is able to extract more information from each choice question. Therefore, it is able to achieve greater predictive performance with fewer questions. It takes approximately 12 choice questions for the best choice-only benchmark to reach the performance achieved by the proposed model after 8 choice questions, and the performance of the best choice-only benchmark after 16 questions is similar to the performance of the proposed model after only 12 questions. In addition, the choice-only benchmark that uses eye-tracking data to account for which cells were visited in each question performs worse than the standard choice-only benchmark that assumes that all cells are visited. Thus, to extract valuable information from eye-tracking data, it is not enough to merely capture which information the consumer processed; it is better to endogenize the information search process. Indeed, simply taking into account the fact that an attribute was ignored in a question without modeling why it was ignored results in no information being collected about this attribute in that question (i.e., the partworths for this attribute do not appear in the likelihood function for that question).

Figure 12 compares the proposed model with the search+choice benchmarks. Again, the proposed model performs significantly better than either of the benchmarks. The worst-performing benchmark is the one that assumes that consumers ignore future product-related utility when deciding whether and how to acquire information. The benchmark that assumes that consumers take future prod-

Figure 12
PROPOSED MODEL VERSUS SEARCH+CHOICE BENCHMARKS: AVERAGE HOLDOUT HIT RATE VERSUS NUMBER OF QUESTIONS USED FOR CALIBRATION

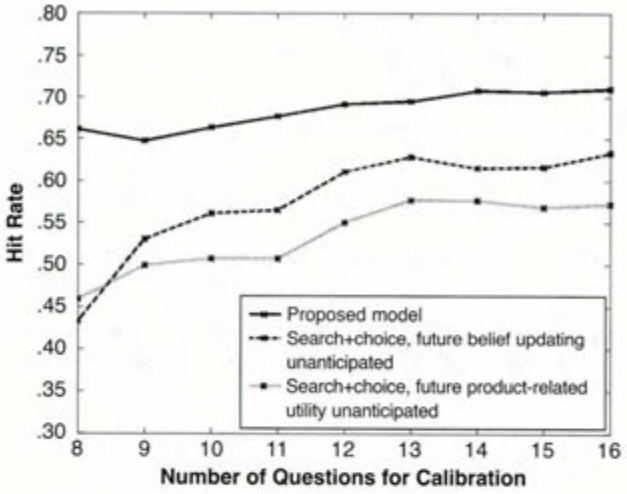
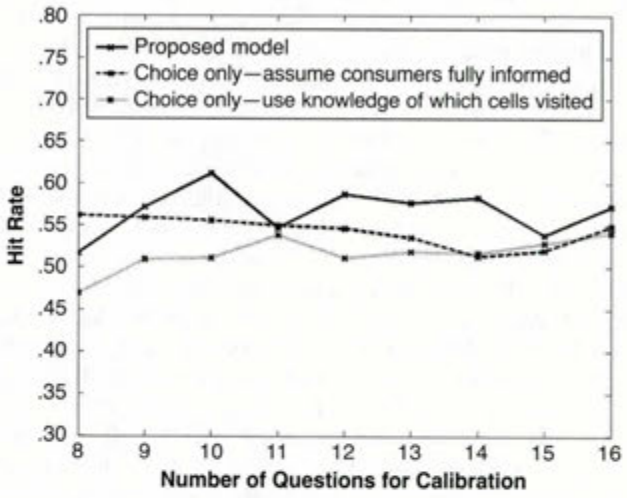


Figure 13
PROPOSED MODEL VERSUS CHOICE-ONLY BENCHMARKS: AVERAGE EXTERNAL VALIDITY HIT RATE VERSUS NUMBER OF QUESTIONS USED FOR CALIBRATION



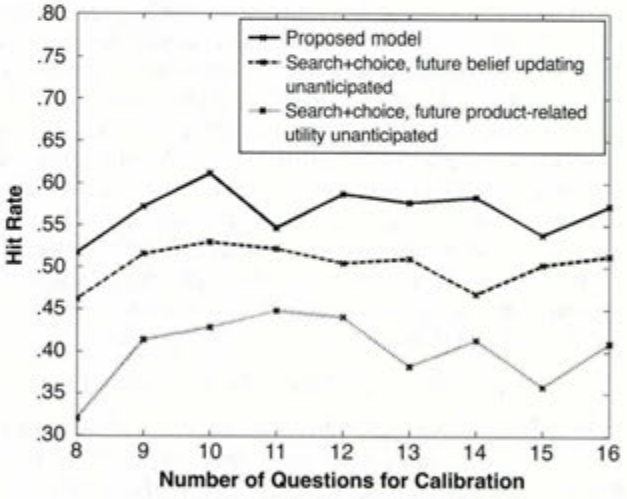
uct-related utility into account but ignore how their beliefs will be updated performs better but still not as well as the proposed model. This finding suggests that the gain from the proposed model comes from assuming that consumers take into account both future utility and the impact that additional search will have on their future decisions.

Figures 13 and 14 compare the models' performance on the external validity task. Although the proposed model still performs better than the benchmarks, and the average difference remains significant (see Tables C1 and C2 in Appendix C), the comparisons are a little noisier for at least two reasons. First, the external validity task was a single choice question, whereas the holdout comparisons are based on the last four questions, which reduces the variance in performance across consumers. Second, the choice shares of the eight alternatives in the external validity question were very unevenly distributed; 95.0% of the respondents chose the three most-popular alternatives (with respective shares of 55.7%, 28.6%, and 11.4%).

CONCLUSIONS

In this article, we develop a joint model of information processing and choice that explicitly captures the dynamic trade-off between search-related utility and product-related utility. We find that the proposed model offers better out-of-sample predictions than benchmarks that either do not leverage data on the information search process or do so without endogenizing search as the outcome of forward-looking utility maximization. Our model also allows for greater discrimination between various attributes than the benchmarks that only model choice and not search. Our results suggest that the gains in predictive performance come from modeling consumers as being forward looking both in terms of taking future utility into account when deciding whether and how to search and in terms of anticipating how search

Figure 14
PROPOSED MODEL VERSUS SEARCH+CHOICE BENCHMARKS: AVERAGE EXTERNAL VALIDITY HIT RATE VERSUS NUMBER OF QUESTIONS USED FOR CALIBRATION



will affect their future beliefs and, therefore, future expected utility.

Our contribution is both methodological and managerial. Methodologically, our model extends Gabaix et al.'s (2006) DC model in several important ways. That model had a single parameter (opportunity cost of time) that was estimated at the aggregate level by matching moments of the data. In contrast, our model specifies a rich search-related utility function that captures fatigue and proximity effects and a product-related utility function parameterized by a set of partworths. Moreover, our model allows consumers to have imperfect memory encoding. We estimate our model

within a likelihood-based hierarchical Bayesian framework that allows for heterogeneity across consumers.

Managerially, as we discussed previously, there are commercial solutions available today that allow for collection of eye-tracking data in an online environment using the consumer's webcam. We expect such solutions to be increasingly common as large companies such as Facebook acquire such capabilities (in 2012, Facebook acquired GazeHawk, a startup that provides webcam eye-tracking services [Protalinski 2012]) and with the development of open-source solutions. Therefore, we believe that the approach developed in this article will be increasingly accessible to market researchers. We show that complementing choice data with eye-tracking data and modeling eye movements as the outcome of forward-looking utility maximization improve out-of-sample performance, enable practitioners and researchers to use shorter questionnaires, and allow greater discrimination between attributes. We envision eye-tracking data being collected systematically in online market research to augment and improve the responses given by consumers.[9]

Finally, we believe that the present research offers several directions for future studies. First, our model can be extended to account for risk aversion, loss aversion, regrets, and other behavioral phenomena (Hauser, Urban, and Weinberg 1993). Second, our model could provide a framework for developing and testing new theories related to information search and choice. Third, additional physiological measures could be collected during preference measurement tasks, such as facial expressions (Teixeira, Wedel, and Pieters 2012). These additional measures could be incorporated into preference measurement models to further improve predictive performance and reduce the required length of questionnaires. Fourth, our bounded rationality framework may be used to shed new light on the impact of incentives in preference measurement. In our framework, consumers trade off the cognitive costs related to information processing with the benefits derived from their choices. Varying the incentives (e.g., the likelihood of each choice being realized) would affect the expected benefits derived from each choice, which should then influence how much information (and possibly which information) consumers process during the task.

APPENDIX A: ILLUSTRATIVE EXAMPLE

We illustrate our state variables and the computation of product-related utility using a simple example. We assume one attribute (I = 1) with three levels (L = 3) and two alternatives per choice question (J = 2). We assume that alternative 1 has attribute 1 at level 1 and alternative 2 has attribute 1 at level 2.

We have

$$I_i^0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix},$$

and the partworths for the first attribute may be represented as

$$\beta_1 = \begin{bmatrix} \beta_{11} \\ \beta_{12} \end{bmatrix} \text{ or } \tilde{\beta}_1 = \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ -\beta_{11} - \beta_{12} \end{bmatrix}.$$

Before the first fixation, the state variables have the following values:

- $p = \varnothing$, and
- $n_{1,1} = n_{1,2} = 0$.

In addition, we have the following:

- $w_{1,1} = w_{1,2} = [⅓, ⅓, ⅓]$,
- The expected product-related utility of alternative 1 is $w_{1,1}\tilde{\beta}_1 = 0$, and
- The expected product-related utility of alternative 2 is $w_{1,2}\tilde{\beta}_1 = 0$.

Suppose that the first fixation is to cell (1, 1) corresponding to attribute 1 of alternative 1. Then, the state variables evolve to

- $p = (1, 1)$,
- $n_{1,1} = 1$, and
- $n_{1,2} = 0$.

Thus, we have the following:

- $w_{1,1} = [\exp(\eta)/\{2 + \exp(\eta)\}, 1/\{2 + \exp(\eta)\}, 1/\{2 + \exp(\eta)\}]$,
- $w_{1,2} = [⅓, ⅓, ⅓]$,
- The expected product-related utility of alternative 1 is now $w_{1,1}\tilde{\beta}_1 = [\exp(\eta) - 1/\{2 + \exp(\eta)]\tilde{\beta}_{11}$, and
- The expected product-related utility of alternative 2 is now $w_{1,2}\tilde{\beta}_1 = 0$.

Suppose that after t = 15, ten fixations have been made to cell (1, 1), five fixations have been made to cell (1, 2), and the last fixation was on cell (1, 2). Then, we have

- $p = (1, 2)$,
- $n_{1,1} = 10$, and
- $n_{1,2} = 5$.

We also have the following:

- $w_{1,1} = [\exp(10\eta)/\{2 + \exp(10\eta)\}, 1/\{2 + \exp(10\eta)\}, 1/\{2 + \exp(10\eta)\}]$,
- $w_{1,2} = [1/\{2 + \exp(5\eta)\}, \exp(5\eta)/\{2 + \exp(5\eta)\}, 1/\{2 + \exp(5\eta)\}]$,
- The expected product-related utility of alternative 1 is now $w_{1,1}\tilde{\beta}_1 = \{[\exp(10\eta) - 1]/[2 + \exp(10\eta)]\}\beta_{11}$, and
- The expected product-related utility of alternative 2 is now $w_{1,2}\tilde{\beta}_1 = \{[\exp(5\eta) - 1]/[2 + \exp(5\eta)]\}\beta_{12}$.

APPENDIX B: SIMULATION

Data Generation

We simulated a situation similar to our experiment. We used the same number of attributes and levels and the same experimental design. We simulated 70 participants completing the first 16 choice questions from the main task. For each participant and each choice question, we simulated the eye movements and the choice on the basis of the learning parameter η, a set of individual-level search-related utility parameters $\theta_n = [\theta_{0n}, \theta_{1n}, \theta_{2n}, \theta_{3n}]$ defined as in Equation 4, and 18 individual-level partworths β_n.

[9]All code used in this article is available upon request.

The learning parameter was set to $\eta = 3$. All individual-level parameters were drawn from a multivariate normal distribution:

$$[\theta_n, \beta_n] \sim N\left([\theta_0, \beta_0], \begin{bmatrix} \Lambda_\theta & 0 \\ 0 & \Lambda_\beta \end{bmatrix}\right),$$

where Λ_θ was a diagonal matrix with $\text{diag}(\Lambda_\theta) = [.1, .01, .1, .1]$, and Λ_β was the identity matrix. Table B1 reports the average values of θ_n and β_n.

Results

We calibrated our proposed model on the simulated data set using the estimation procedure described in the "Identification and Estimation" subsection. We performed a grid search on the parameter η by calibrating the model with $\eta = 0$ to 5 with increments of 1. The true value $\eta = 3$ was accurately selected on the basis of the DIC.

Table B1 reports the average estimates of the relevant parameters and the 95% credible intervals. We observe that all the search-related parameters as well as 16 of 18 partworths are contained within the 95% credible intervals. The two partworths that fall outside the 95% credible interval (1.9 GHz and one year of Dell support) are still reasonably well recovered.

APPENDIX C: COMPARISONS BETWEEN PROPOSED MODEL AND BENCHMARKS

To compare the performance of the various models statistically, we run ordinary least squares regressions with the

Table B1
SIMULATION RESULTS

	True Average Value	*Estimated Average Value*	*95% Credible Interval*
Search-Related Parameters			
θ_0	1.99	1.94	[1.79, 2.06]
θ_1	−.07	−.07	[−.09, −.06]
θ_2	−.97	−.99	[−1.02, −.97]
θ_3	.70	.74	[.70, .79]
Processor Speed			
1.6 GHz	−2.98	−3.28	[−3.87, −2.65]
1.9 GHz	−1.03	−.63	[−.99, −.33]
2.7 GHz	1.08	.80	[.30, 1.29]
Screen Size			
26 cm	−3.05	−2.95	[−3.41, −2.63]
35.6 cm	−1.14	−1.44	[−1.79, −1.04]
40 cm	.80	1.06	[.65, 1.45]
Hard Drive			
160 GB	−2.83	−3.09	[−3.47, −2.78]
320 GB	−1.02	−.67	[−1.09, −.32]
500 GB	1.02	1.20	[.90, 1.43]
Dell Support			
1 year	−2.85	−3.28	[−3.65, −2.94]
2 years	−.97	−1.03	[−1.23, −.80]
3 years	1.06	1.10	[.79, 1.36]
Antivirus			
30 days	−3.03	−3.05	[−3.44, −2.69]
1 year	−1.12	−1.06	[−1.33, −.73]
2 years	1.08	1.01	[.60, 1.57]
Price			
350€	−3.10	−2.86	[−3.39, −2.32]
500€	−.96	−1.34	[−1.82, −.86]
650€	1.12	1.33	[1.07, 1.56]

Table C1
PROPOSED MODEL VERSUS CHOICE-ONLY BENCHMARKS: REGRESSION RESULTS

	Holdout Questions		*External Validity*	
	Coefficient	p-*Value*	*Coefficient*	p-*Value*
Intercept	.918	.000	.618	.000
Choice only—assume consumers fully informed dummy	−.187	.000	−.518	.002
Choice only—use knowledge of which cells visited dummy	−.426	.000	−.679	.000
q	.040	.000	.006	.843
q^2	.001	.728	−.017	.184
Choice only—assume consumers fully informed dummy × q	.098	.000	−.027	.521
Choice only—use knowledge of which cells visited dummy × q	.088	.000	.048	.259
Choice only—assume consumers fully informed dummy × q^2	−.025	.000	.018	.342
Choice only—use knowledge of which cells visited dummy × q^2	−.018	.001	.009	.645

Notes: The variable q is the (mean-centered) number of questions used for calibration.

Table C2
PROPOSED MODEL VERSUS SEARCH+CHOICE BENCHMARKS: REGRESSION RESULTS

	Holdout Questions		*External Validity*	
	Coefficient	p-*Value*	*Coefficient*	p-*Value*
Intercept	.918	.000	.618	.000
Search+choice—future belief updating unanticipated dummy	−.456	.000	−.731	.000
Search+choice—future product-related utility unanticipated dummy	−.716	.000	−1.342	.000
q	.040	.000	.006	.835
q^2	.001	.728	−.017	.162
Search+choice—future belief updating unanticipated dummy × q	.054	.000	.012	.762
Search+choice—future product-related utility unanticipated dummy × q	.024	.053	−.007	.861
Search+choice—future belief updating unanticipated dummy × q^2	−.020	.000	.011	.551
Search+choice—future product-related utility unanticipated dummy × q^2	−.010	.076	−.002	.888

Notes: The variable q is the (mean-centered) number of questions used for calibration.

logit of the hit rate as the dependent variable.[10] The number of observations in each regression is the number of respondents × 9 (number of questions used for calibration varies from 8 to 16) × 3 (number of models being compared). We include respondent fixed effects to capture the panel structure of the data. We use the proposed model as baseline and include dummy variables for benchmark models. We also include covariates that capture the increasing trend in performance as the number of questions increases. The benchmarks in Table C1 (Table C2) are the choice-only (search+choice) models. We run one set of regressions for the holdout questions and one set for the external validity task. We find that all main effects corresponding to the benchmarks are significantly negative; the proposed model on average performs significantly better than all benchmarks.

REFERENCES

Allenby, Greg M., Neeraj Arora, and James L. Ginter (1995), "Incorporating Prior Knowledge into the Analysis of Conjoint Studies," *Journal of Marketing Research*, 32 (May), 152–62.

Assunção, João L. and Robert J. Meyer (1993), "The Rational Effect of Price Promotions on Sales and Consumption," *Management Science*, 39 (5), 517–35.

Atchadé, Yves F. and Jeffrey S. Rosenthal (2005), "On Adaptive Markov Chain Monte Carlo Algorithms," *Bernoulli*, 11 (5), 815–28.

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong (2004), "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics*, 119 (3), 861–98.

Celeux, Gilles, Florence Forbes, Christian P. Robert, and D. Michael Titterington (2006), "Deviance Information Criteria for Missing Data Models," *Bayesian Analysis*, 1 (4), 651–73.

Chandon, Pierre, J. Wesley Hutchinson, Eric T. Bradlow, and Scott H. Young (2009), "Does In-Store Marketing Work? Effects of the Number and Position of Shelf Facings on Brand Attention and Evaluation at the Point of Purchase," *Journal of Marketing*, 73 (November), 1–17.

Ching, Andrew T., Susumu Imai, Masakazu Ishihara, and Neelam Jain (2012), "A Practitioner's Guide to Bayesian Estimation of Discrete Choice Dynamic Programming Models," *Quantitative Marketing and Economics*, 10 (2), 1–46.

Chintagunta, Pradeep K., Ronald L. Goettler, and Minki Kim (2012), "New Drug Diffusion When Forward-Looking Physicians Learn from Patient Feedback and Detailing," *Journal of Marketing Research*, 49 (December), 807–821.

Ding, Min (2007), "An Incentive-Aligned Mechanism for Conjoint Analysis," *Journal of Marketing Research*, 44 (May), 214–23.

——, Rajdeep Grewal, and John Liechty (2005), "Incentive-Aligned Conjoint Analysis," *Journal of Marketing Research*, 42 (February), 67–82.

Dubé, Jean-Pierre H., Günter J. Hitsch, and Pranav Jindal (2012), "The Joint Identification of Utility and Discount Functions from Stated Choice Data: An Application to Durable Goods Adoption," working paper, National Bureau of Economic Research.

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg (2006), "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model," *American Economic Review*, 96 (4), 1043–68.

Gelman, Andrew, Xiao-Li Meng, and Hal Stern (1996), "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies," *Statistica Sinica*, 6 (4), 733–59.

Gilbride, Timothy J. and Greg M. Allenby (2004), "A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules," *Marketing Science*, 23 (3), 391–406.

—— and —— (2006), "Estimating Heterogeneous EBA and Economic Screening Rule Choice Models," *Marketing Science*, 25 (5), 494–509.

Gittins, John C. (1979), "Bandit Processes and Dynamic Allocation Indices," *Journal of the Royal Statistical Society, Series B (Methodological)*, 41 (2), 148–77.

Hagerty, Michael R. and David A. Aaker (1984), "A Normative Model of Consumer Information Processing," *Marketing Science*, 3 (3), 227–46.

Hartmann, Wesley R. and Harikesh S. Nair (2010), "Retail Competition and the Dynamics of Demand for Tied Goods," *Marketing Science*, 29 (2), 366–86.

Hauser, John R., Glen L. Urban, and Bruce D. Weinberg (1993), "How Consumers Allocate Their Time When Searching for Information," *Journal of Marketing Research*, 30 (November), 452–66.

Huang, Guofang, Ahmed Khwaja, and K. Sudhir (2012), "Short Run Needs and Long Term Goals: A Dynamic Model of Thirst Management," working paper, Cowles Foundation for Research in Economics, Yale University.

Hutchinson, J. Wesley and Robert J. Meyer (1994), "Dynamic Decision Making: Optimal Policies and Actual Behavior in Sequential Choice Problems," *Marketing Letters*, 5 (4), 369–82.

Jedidi, Kamel and Rajeev Kohli (2005), "Probabilistic Subset-Conjunctive Models for Heterogeneous Consumers," *Journal of Marketing Research*, 42 (November), 483–94.

—— and —— (2008), "Inferring Latent Class Lexicographic Rules from Choice Data," *Journal of Mathematical Psychology*, 52 (4), 241–49.

Johnson, Eric J., Robert J. Meyer, and Sanjoy Ghose (1989), "When Choice Models Fail: Compensatory Models in Negatively Correlated Environments," *Journal of Marketing Research*, 26 (August), 255–70.

Liechty, John, Rik Pieters, and Michel Wedel (2003), "Global and Local Covert Visual Attention: Evidence from a Bayesian Hidden Markov Model," *Psychometrika*, 68 (4), 519–41.

Meyer, Robert J. (1982), "A Descriptive Model of Consumer Information Search Behavior," *Marketing Science*, 1 (1), 93–121.

Misra, Sanjog and Harikesh S. Nair (2011), "A Structural Model of Sales-Force Compensation Dynamics: Estimation and Field Implementation," *Quantitative Marketing and Economics*, 9 (3), 211–57.

Musalem, Andres, Martin Meißner, and Joel Huber (2013), "Do Motivated and Incidental Processing Distort Conjoint Choices?" working paper, Duke University.

Payne, John W., James R. Bettman, and Eric J. Johnson (1988), "Adaptive Strategy Selection in Decision Making," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14 (3), 534–52.

——, ——, and —— (1992), "Behavioral Decision Research: A Constructive Processing Perspective," *Annual Review of Psychology*, 43 (1), 87–131.

——, ——, and —— (1993), *The Adaptive Decision Maker*. Cambridge, UK: Cambridge University Press.

Pieters, Rik, Edward Rosbergen, and Michel Wedel (1999), "Visual Attention to Repeated Print Advertising: A Test of Scanpath Theory," *Journal of Marketing Research*, 36 (November), 424–38.

—— and Luk Warlop (1999), "Visual Attention During Brand Choice: The Impact of Time Pressure and Task Motivation," *International Journal of Research in Marketing*, 16 (1), 1–16.

[10]We take the logit because the hit rate is bounded between 0 and 1.

——, ——, and Michel Wedel (2002), "Breaking Through the Clutter: Benefits of Advertisement Originality and Familiarity for Brand Attention and Memory," *Management Science*, 48 (6), 765–81.

—— and Michel Wedel (2004), "Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects," *Journal of Marketing*, 68 (April), 36–50.

Protalinski, Emil (2012), "Facebook Acquires Eye-Tracking Startup GazeHawk," ZDNet.com, (March 8), (accessed January 9, 2015), [available at www.zdnet.com/blog/ facebook/facebook-acquires-eye-tracking-startup-gazehawk/10218].

Rosbergen, Edward, Rik Pieters, and Michel Wedel (1997), "Visual Attention to Advertising: A Segment-Level Analysis," *Journal of Consumer Research*, 24 (3), 305–314.

Rust, John (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55 (5), 999–1033.

Shi, Savannah Wei, Michel Wedel, and F.G.M. (Rik) Pieters (2013), "Information Acquisition During Online Decision Making: A Model-Based Exploration Using Eye-Tracking Data," *Management Science*, 59 (5), 1009–1026.

Simon, Herbert A. (1955), "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, 69 (1), 99–118.

Stüttgen, Peter, Peter Boatwright, and Robert T. Monroe (2012), "A Satisficing Choice Model," *Marketing Science*, 31 (6), 878–99.

Teixeira, Thales, Michel Wedel, and Rik Pieters (2012), "Emotion-Induced Engagement in Internet Video Advertisements," *Journal of Marketing Research*, 49 (April), 144–59.

Toubia, Olivier, Martijn G. de Jong, Daniel Stieger, and Johann Füller (2012), "Measuring Consumer Preferences Using Conjoint Poker," *Marketing Science*, 31 (1), 138–56.

—— and Andrew T. Stephen (2013), "Intrinsic vs. Image-Related Utility in Social Media: Why Do People Contribute Content to Twitter?" *Marketing Science*, 32 (3), 368–92.

Tversky, Amos, Shmuel Sattath, and Paul Slovic (1988), "Contingent Weighting in Judgment and Choice," *Psychological Review*, 95 (3), 371–84.

Van der Lans, Ralf, Rik Pieters, and Michel Wedel (2008a), "Competitive Brand Salience," *Marketing Science*, 27 (5), 922–31.

——, ——, and —— (2008b), "Eye-Movement Analysis of Search Effectiveness," *Journal of the American Statistical Association*, 103 (482), 452–61.

——, Michel Wedel, and Rik Pieters (2011), "Defining Eye-Fixation Sequences Across Individuals and Tasks: The Binocular-Individual Threshold (BIT) Algorithm," *Behavior Research Methods*, 43 (1), 239–57.

Wedel, Michel and Rik Pieters (2000), "Eye Fixations on Advertisements and Memory for Brands: A Model and Findings," *Marketing Science*, 19 (4), 297–312.

—— and —— (2008), *Eye Tracking for Visual Marketing*. Delft, The Netherlands: Now Publishers.

——, ——, and John Liechty (2008), "Attention Switching During Scene Perception: How Goals Influence the Time Course of Eye Movements Across Advertisements," *Journal of Experimental Psychology: Applied*, 14 (2), 129–38.

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

Yao, Song and Carl F. Mela (2011), "A Dynamic Model of Sponsored Search Advertising," *Marketing Science*, 30 (3), 447–68.

Yee, Michael, Ely Dahan, John R. Hauser, and James Orlin (2007), "Greedoid-Based Noncompensatory Inference," *Marketing Science*, 26 (4), 532–49.