

Website Morphing

John R. Hauser, Glen L. Urban

MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142
{hauser@mit.edu, glurban@mit.edu}

Guilherme Liberali

MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, and
Universidade do Vale do Rio dos Sinos, Sao Leopoldo, RS 90450 Brazil, liberali@unisinos.br

Michael Braun

MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142,
braunm@mit.edu

Virtual advisors often increase sales for those customers who find such online advice to be convenient and helpful. However, other customers take a more active role in their purchase decisions and prefer more detailed data. In general, we expect that websites are more preferred and increase sales if their characteristics (e.g., more detailed data) match customers' cognitive styles (e.g., more analytic). "Morphing" involves automatically matching the basic "look and feel" of a website, not just the content, to cognitive styles. We infer cognitive styles from clickstream data with Bayesian updating. We then balance exploration (learning how morphing affects purchase probabilities) with exploitation (maximizing short-term sales) by solving a dynamic program (partially observable Markov decision process). The solution is made feasible in real time with expected Gittins indices. We apply the Bayesian updating and dynamic programming to an experimental BT Group (formerly British Telecom) website using data from 835 priming respondents. If we had perfect information on cognitive styles, the optimal "morph" assignments would increase purchase intentions by 21%. When cognitive styles are partially observable, dynamic programming does almost as well—purchase intentions can increase by almost 20%. If implemented system-wide, such increases represent approximately \$80 million in additional revenue.

Key words: Internet marketing; cognitive styles; dynamic programming; Bayesian methods; clickstream analysis; automated marketing; website design; telecommunications

History: Received: December 7, 2007; accepted August 25, 2008; processed by Fred Feinberg. Published online in *Articles in Advance* February 24, 2009.

1. Introduction and Motivation

Website design has become a major driver of profit. Websites that match the preferences and information needs of visitors are efficient; those that do not forego potential profit and may be driven from the market. For example, when Intel redesigned its website by adding a verbal advisor to help customers find the best software to download for their digital cameras, successful downloads increased by 27%.¹ But verbal advisors are not for every customer. Less verbal and more analytic customers found the verbal advisor annoying and preferred a more graphic list of downloadable software. If customers vary in the way they process information (that is, vary in their cognitive styles), Intel might increase downloads even

more with a website that automatically changes its characteristics to match those cognitive styles.

Intel is not alone. Banks, cell phone providers, broadband providers, content providers, and many retailers might serve their customers better and sell more products and services if their websites matched the cognitive styles of their visitors. One solution is personalized self-selection, in which a customer is given many options and allowed to select how to navigate and interact with the site. As the customer's options grow, this strategy leads to sites that are complex, confusing, and difficult to use. Another option, popular in the adaptive-learning literature, is to require visitors to complete a set of cognitive-style tasks and then select a website from a predetermined set of websites. However, retail website visitors are likely to find such intensive measurement cumbersome and intrusive. They may leave the website before completing such tasks.

We propose another approach: "morphing" the website automatically by matching website characteristics to customers' cognitive styles. Our practical

¹ Although downloads are free, the benefits to Intel are substantial in terms of enhanced customer satisfaction, increased sales of hardware, and cost savings because of fewer telephone-support calls. The cost savings alone saved Intel over \$1 million for their camera products with an estimated \$30 million in savings across all product categories (Rhoads et al. 2004). Figure 1 illustrates one virtual advisor. See Urban and Hauser (2004) for other examples.

goal is to morph the website's basic structure (site backbone) and other functional characteristics in real time. Website morphing complements self-selected branching (as in <http://www.Dell.com>), recommendations (as in <http://www.Amazon.com>), factorial experiments (Google's Website Optimizer), or customized content (Ansari and Mela 2003, Montgomery et al. 2004). Website morphing is an example of targeting optimal marketing communications to customer segments (Tybout and Hauser 1981, Wernerfelt 1996).

Example dimensions on which cognitive styles are measured might include impulsive (makes decisions quickly) versus deliberative (explores options in depth before making a decision), visual (prefers images) versus verbal (prefers text and numbers), or analytic (wants all details) versus holistic (just the bottom line). (We provide greater detail and citations in §7.) A website might morph by changing the ratio of graphs and pictures to text, by reducing a display to just a few options (broadband service plans), or by carefully selecting the amount of information presented about each plan. A website might also morph by adding or deleting functional characteristics such as column headings, links, tools, persona, and dialogue boxes.

Website morphing presents at least four technical challenges. (1) For first-time visitors, a website must morph based on relatively few clicks; otherwise, the customer sees little benefit. (2) Even if we knew a customer's cognitive style, the website must learn which characteristics are best for which customers (in terms of sales or profit). (3) To be practical, a system needs prior distributions on parameters. (4) Implementation requires a real-time working system (and the inherently difficult Web programming).

We use a Bayesian learning system to address the rapid assessment of cognitive styles and a dynamic program to optimally manage the tension between exploitation (serving the morph most likely to be best for a customer) and exploration (serving alternative morphs to learn which morph is best). Uncertainty in customer styles implies a partially observable Markov decision process (POMDP), which we address with fast heuristics that are close to optimal. Surveys, using both conjoint analysis and experimentation, provide priors and "prime" the Bayesian and dynamic programming engines. We demonstrate feasibility and potential profit increases with an experimental website developed for the BT Group to sell broadband service in Great Britain.

2. An Adaptive System to Infer Cognitive Styles and Identify Optimal Morphs

A cognitive style is "a person's preferred way of gathering, processing, and evaluating information"

(Hayes and Allinson 1998, p. 850) and can be identified as "individual differences in how we perceive, think, solve problems, learn and relate to others" (Witkin et al. 1977, p. 15). "A person's cognitive style is...fixed early on in life and is thought to be deeply pervasive...[and is] a relatively fixed aspect of learning performance" (Riding and Rayner 1998, p. 7). Cognitive styles tend to be forced-choice (ipsative) constructs, such as analytic versus holistic, and are usually measured by question banks or cognitive tasks (Frias-Martinez et al. 2007, Santally and Alain 2006, Riding and Rayner 1998).

The literature is wide and varied. We derive a flexible system that works with any reasonable set of cognitive-style dimensions. We illustrate the system with commonly used cognitive-style constructs found in the literature (§7, BT application).

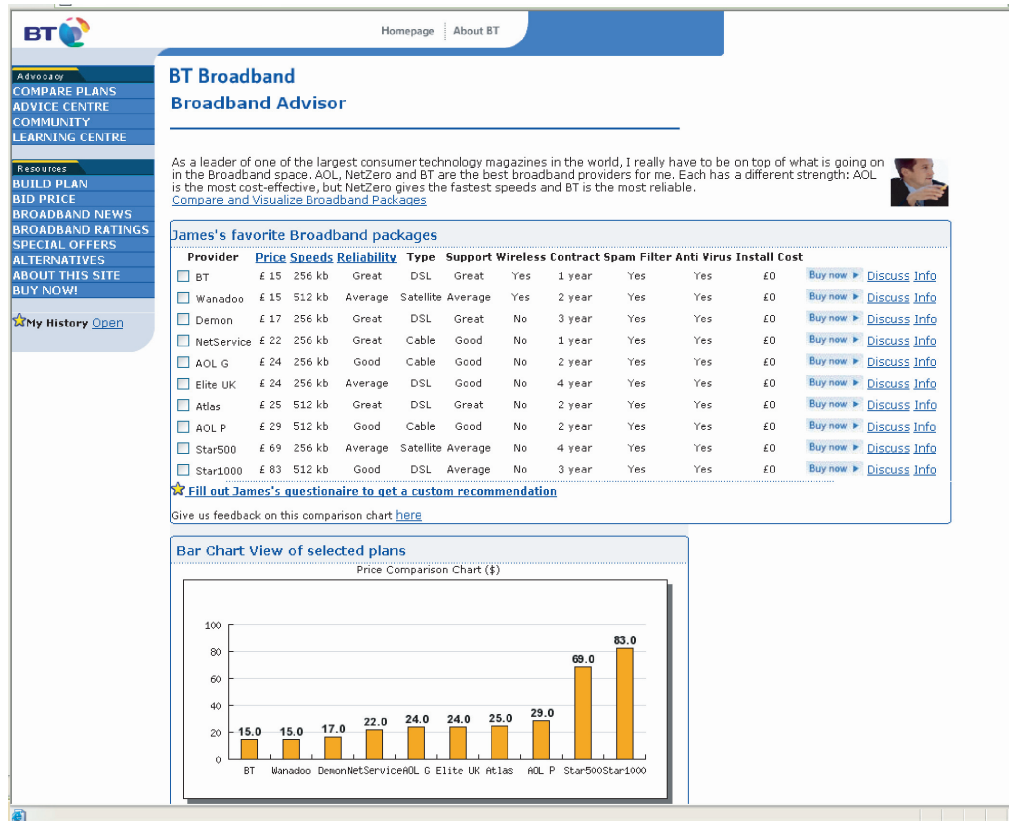
Figure 1 illustrates two of the eight versions ("morphs") of broadband advisors from the BT application. Figure 1(a) uses an analytic virtual advisor (a technology magazine editor willing to provide data) who compares plans on 10 characteristics (a large information load), presents a bar chart to compare prices (graphical), and provides technical information about plans (focused content). In contrast, Figure 1(b) uses an holistic virtual advisor (typical user) to whom the website visitor can listen (verbal). This advisor avoids details, compares plans on only four characteristics (small information load), and gives an easy-to-comprehend overall comparison of three plans (general content).

We expect different morphs to appeal differentially depending on visitors' cognitive styles. For example, impulsive visitors might prefer less-detailed information, whereas deliberative visitors might prefer more information. Similarly, the more focused of the two morphs might appeal to visitors who are holistic, while the ability to compare many plans in a table might appeal to analytic visitors. If preferences match behavior (an empirical question), then, by matching a website's characteristics to cognitive styles, the morphing website should sell broadband service more effectively and lead to greater profits for BT.

We defer to §7 the selection, definition, and measurement of cognitive styles, the definition and implementation of website characteristics (morphs), and the market research that provides prior beliefs (purchase probabilities) on the relationships between cognitive styles and morph characteristics. For BT we use four binary cognitive-style constructs yielding $2^4 = 16$ cognitive-style segments, indexed by r_n for the n th website visitor (customer). We attempt to morph the BT website to match cognitive styles of each segment by using three binary website characteristics yielding $2^3 = 8$ possible morphs, indexed by m .

Figure 1 Comparison of Two Morphs for a Website Advisor

(a) Focused content, large-load, graphical morph



(b) General content, small-load, verbal morph

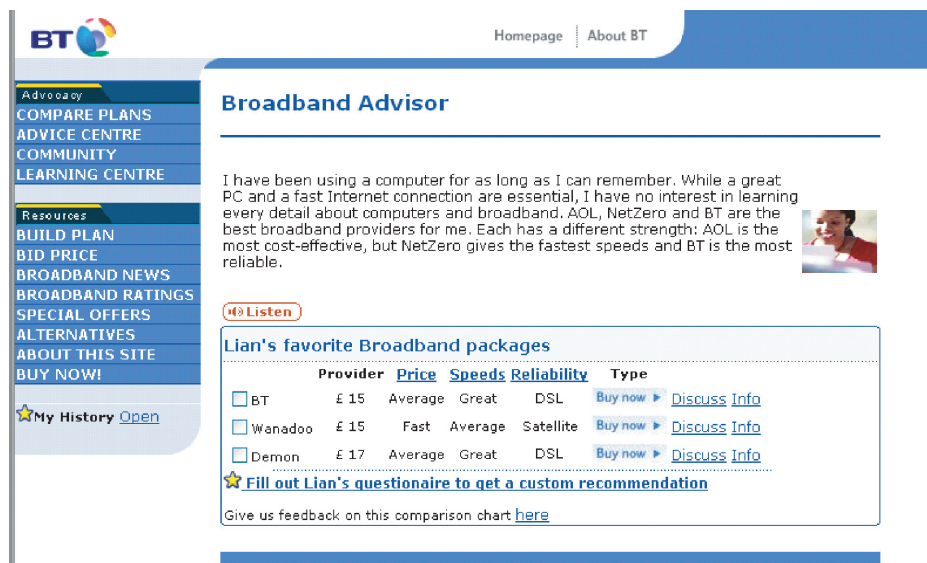
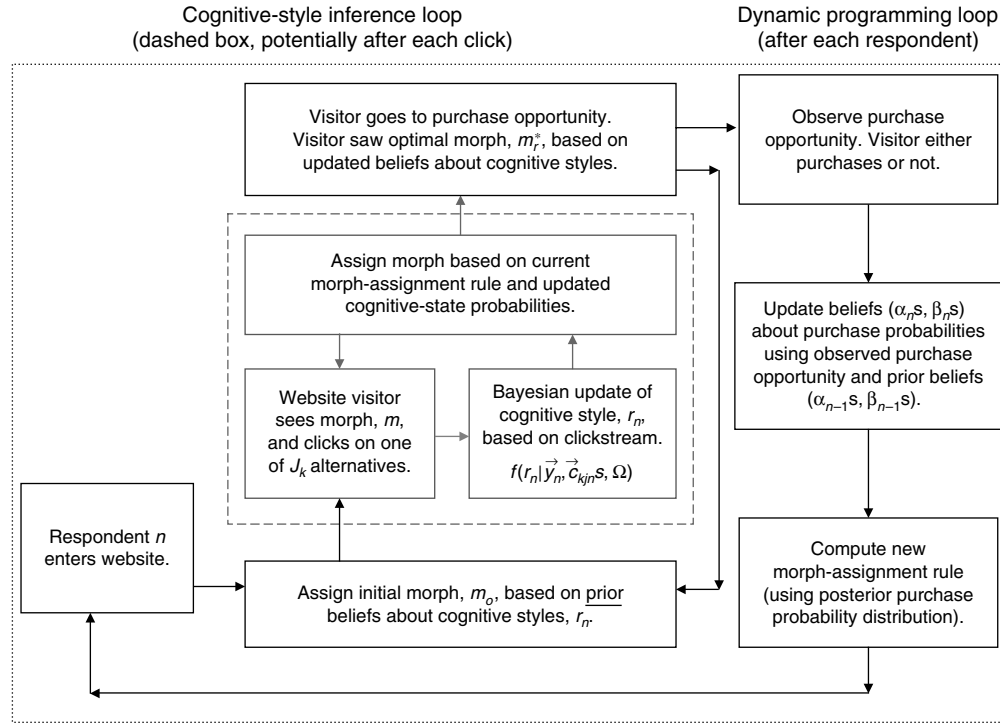


Figure 2 Website Morphing: Bayesian Inference Loop (Grey Dashed Line) and Dynamic Programming Inference/Optimization Loop (Black Dotted Line)



If we had perfect information on cognitive-style segments and perfect knowledge of segment $\times$ morph purchase probabilities, we could map an optimal morph to each cognitive-style segment. There are $16 \times 8 = 128$ such segment $\times$ morph probabilities. In the absence of perfect information, our challenge is to infer the cognitive-segment to which each visitor belongs while simultaneously learning how to maximize profit by assigning morphs to cognitive-style segments.

In real systems, we must infer visitors' cognitive-style segment from their clickstreams. We can do this because each visitor's click is a decision point that reveals the visitor's cognitive-style preferences. If we observe a large number of clicks, we should be able to identify a visitor's cognitive-style segment well. However, in any real application, the number of clicks we observe before morphing will be relatively small, yielding at best a noisy indicator of segment membership.

The website begins with morph m_o (to be determined). We observe some number of clicks (say, 10), infer probabilities for the visitor's cognitive-style segment, then morph the website based on our inference of the visitor's segment. The visitor continues until he or she purchases (a broadband service) or exits the website without purchasing. In our application, maximizing purchases is a good surrogate for maximizing profit through the Web channel. (In §11 we indicate

how to extend our framework to address the size of the purchase.)

We begin with the Bayesian inference loop (grey dashed line in Figure 2) through which we infer the visitor's cognitive-style segment. Denote by J_{kn} the number of potential click-alternatives that the n th visitor faces on the k th click. Let $y_{k:n}$ be 1 if the n th visitor chooses the j th alternative on the k th click, and 0 otherwise. Let $\vec{y}_{k:n}$ be the vector of the $y_{k:n}$ s and let $\vec{y}_n$ be the matrix of the $\vec{y}_{k:n}$ s. Each click-alternative is described by a set of characteristics, $\vec{c}_{k:n}$. In our application, there are 11 characteristics: three macro characteristics (e.g., visual versus verbal), four detailed function characteristics (e.g., a link that plays audio), and four topical website areas (e.g., virtual advisor). All notation is summarized in Appendix 1 for easy reference.

A visitor in a particular cognitive-style segment will prefer some combinations of characteristics to other combinations. Let $\vec{\omega}_{r_n}$ be a vector of preference weights that maps click-alternative characteristics, $\vec{c}_{k:n}$, to preferences for each cognitive-style segment, r_n . Define Ω as the matrix of the $\vec{\omega}_{r_n}$ s. If we know (1) preferences for morph characteristics for each cognitive-style segment, (2) morph characteristics for click-alternatives (various links on which the visitor can click when he or she makes a decision to click), and (3) the clicks that were made, we can infer the visitor's cognitive-style segment using

Bayes' theorem. Specifically, we update the posterior distribution, $f(r_n | \tilde{y}_n, \Omega, \tilde{c}_{k|n}s)$, that the visitor is in the r_n^{th} segment based on the observed data.²

The second inference loop (outer loop denoted by a black dotted line in Figure 2) identifies the optimal morph conditioned on $f(r_n | \tilde{y}_n, \Omega, \tilde{c}_{k|n}s)$. This inference loop must learn and optimize simultaneously. In theory, we might allow the website to morph many times for each visitor, potentially after every click. However, in our application we observe only one purchase decision per visitor. To avoid unnecessary assumptions in assigning this purchase to website characteristics, our initial application morphs only once per visitor. (We address alternative strategies in §5.) Any results we report are conservative and might be improved with future websites that morph more often (potentially taking switching costs, if any, into account).

Let p_{rm} be the probability that a visitor in cognitive-style segment, $r_n = r$, will purchase BT's broadband plan after visiting a website that has the characteristics of morph m . Let $\tilde{p}$ be the matrix of the p_{rm} s. Clearly, if we knew r_n and the $\tilde{p}$ perfectly, then we would assign the morph that maximizes p_{rm} . However, we do not know either r_n or $\tilde{p}$ perfectly; we have only posterior probabilistic beliefs about r_n and $\tilde{p}$. Without perfect information, maximizing long-term expected profit (sales) requires that we solve a much more difficult problem.

For example, suppose we knew r_n but had only posterior beliefs about p_{rm} . A naïve myopic strategy might choose the morph m , which has the largest (posterior) mean for p_{rm} . But the naïve strategy does not maximize long-term profits. There might be another morph, m' , with a lower (posterior) mean but with a higher variance in (posterior) beliefs. We might choose m' to sacrifice current profits but learn more about the distribution of $p_{rm'}$. The knowledge gained might help us make better decisions in the future. We are more likely to choose m' when we value future decisions and when we benefit greatly from reducing the uncertainty in $p_{rm'}$. The optimal morph-assignment problem is even more difficult when we face uncertainty about the cognitive-style segment, r_n . We must also take into account "false negatives" when we assign a morph that is not right for the true cognitive-style segment. This is an explicit opportunity cost to BT for which we must account when we assign morphs to maximize profit.

To maximize profit, taking both exploration and potential false negatives into account, we formulate a dynamic program. When r is known, the solution

is based on a well-studied structure ("multiarmed bandits"). The optimal morph-assignment rule can be computed between clicks to automatically balance exploration and exploitation. When r is unknown, the partial-information optimal solution is not feasible between clicks. Instead, we use a fast heuristic that obtains 99% of long-term profits (sales) when all uncertainty is taken into account. (We test both dynamic programming solutions on our data.)

Before we formulate these dynamic programs we briefly review prior attempts to adapt content to latent characteristics of users of that content.

3. Related Prior Literature

Cognitive styles (also learning styles or knowledge levels) have been used to adapt material for distance learning, Web-based learning, digital libraries, and hypermedia navigation. In most cases, cognitive styles are measured with an intensive inventory of psychometric scales or inferred from predefined tasks (Frias-Martinez et al. 2007, Magoulas et al. 2001, Mainemelis et al. 2002, Santally and Alain 2006, Tarpin-Bernard and Habieb-Mammar 2005). Methods include direct classification, neuro-fuzzy logic, decision trees, multilayer perceptrons, Bayesian networks, and judgment. Most authors match the learning or search environment based on judgment by an expert pedagogue or based on predefined distance measures. In contrast we infer cognitive styles from relatively few clicks and automatically balance exploration and exploitation to select the best morph.

Automatic assignment is common in statistical machine learning. For example, Chickering and Paek (2007) use reinforcement learning to infer a user's commands from spoken language. After training the system with 20,000 synthetic voices, they demonstrate that the system becomes highly accurate after 1,000 spoken commands. Like us, they formulate their problem as a multiarmed bandit, but their focus and data require an entirely different solution strategy.

When latent customer states are transient, hidden Markov models (HMMs) have proven useful. Conati et al. (2002) identify students' mastery of Newton's laws by predefining a Bayesian network and updating hidden-state probabilities by observing students' answers. Conditional probabilities are set by judgment. Their intelligent tutoring system (ITS) provides hints for "rules" when it infers that a student has not yet mastered the lesson. Yudelson et al. (2008) extend this ITS with more hidden states and estimate the parameters of the Bayesian network with an expectation-maximization algorithm. In other HMM models, Bidel et al. (2003) identify navigation strategies for hypermedia, Liechty et al. (2003) identify visual attention levels for advertising,

² This posterior distribution depends on the morph, m_o , that the n th visitor has experienced. We have suppressed this subscript for ease of exposition. We explore estimation of Ω in §8.

and Netzer et al. (2008) identify customer attitudes for alumni gift giving. Estimation methods include machine learning and hierarchical Bayes Monte Carlo Markov chain methods. Montoya et al. (2008) estimate an HMM and optimize sampling and detailing with dynamic programming.

HMMs have proven accurate in these situations and policy simulations suggest significant profit increases. However, HMMs are computationally intensive, often requiring more than a day of computer time to estimate parameters and almost as long to optimize policies. In contrast, we compute strategies in real time between clicks (Bayesian inference loop) and update strategies between online visitors (dynamic programming loop). Because we expect cognitive styles to be enduring characteristics of website visitors (e.g., Riding and Raynor 1998), we avoid the computational demands necessary to model transient latent states. In our application we use priming data and ipsative scales to identify cognitive style segments (see §7 and the Technical Appendix, available at <http://mktsci.pubs.informs.org>, on morphing taxonomies). Alternatively, one might consider latent-class analyses to uncover enduring cognitive-style segments.

We now present a working system in which we combine and adapt known methods to website morphing.

4. Finding the Optimal Morph with Gittins Indices

We present the dynamic programming solution in steps. In this section we *temporarily* assume that the visitor sees morph m for the entire visit and we know the visitor's cognitive segment, r . In the next section we relax these assumptions to solve a partially observable Markov decision process where we infer r and where the visitor may not see morph m for the entire visit.

Let $\delta_{mn} = 1$ if the n th visitor purchases a BT broadband plan after seeing morph, m . Let $\delta_{mn} = 0$ otherwise. For clarity of exposition when r is known, we write δ_{mn} as δ_{rmn} to make the dependence on r explicit. Under the temporary assumption that r is known, we model the observed broadband subscriptions, δ_{rmn} , as outcomes of a Bernoulli process with probability, p_{rm} . Based on these purchase observations and prior beliefs, we infer a posterior distribution on purchase probabilities, $f(\bar{p} | \delta_{rmn}, \text{parameters based on previous visitors})$.

To represent our prior beliefs, we choose a flexible family of probability distributions that is naturally conjugate to the Bernoulli process. The conjugate prior is a beta distribution with morph-and-segment-specific parameters α_{rmo} and β_{rmo} . Specifically,

$$f(p_{rm} | \alpha_{rmo}, \beta_{rmo}) \sim p_{rm}^{\alpha_{rmo}-1} (1 - p_{rm})^{\beta_{rmo}-1}.$$

With beta priors and Bernoulli observations, it is easy to show that the posterior is also a beta distribution with $\alpha_{rm,n+1} = \alpha_{rmn} + \delta_{rmn}$ and $\beta_{rm,n+1} = \beta_{rmn} + (1 - \delta_{rmn})$. If a visitor in segment r receives morph m , we expect an immediate expected reward equal to the mean of the beta distribution, $\bar{p}_{rmn} = \alpha_{rmn} / (\alpha_{rmn} + \beta_{rmn})$, times the profit BT earns if the n th visitor purchases a broadband plan. We also earn an expected reward for acting optimally in the future, which we discount by a . The solution to the dynamic program is the morph, m_r^* , which maximizes the sum of the expectation of the immediate reward and the discounted future reward.

In general, such multiarm bandit dynamic programs are difficult to solve. In fact, “during the Second World War [this problem was] recognized as so difficult that it quickly became... a by-word for intransigence” (Whittle 1989, p. ix). However, in a now-classic paper, Gittins (1979) proposed a simple and practical solution that decomposed the problem into indices. In the Gittins solution a candidate “arm,” in our case a morph, is compared to an arm for which the payoff probability is known with certainty. Gittins formulates the Bellman equation (given below) and solves for this known payoff probability, which we denote by G_{rmn} . G_{rmn} depends only on α_{rmn} , β_{rmn} , and a and is independent of the parameters of the other arms. This known payoff probability has become known as the Gittins index. Gittins proved that these indices contain all of the information necessary to select the optimal strategy at any point in time, automatically balancing exploitation and exploration. Gittins' solution is simply to choose the arm with the largest index.³ Future morph assignments might change when we update α_{rmn+1} , β_{rmn+1} , and G_{rmn+1} with new information. However, the strategy of choosing the highest-index morph is always optimal.

Gittins (1979) proof of indexability is beyond the scope of this paper. However, it is instructive to formulate the Bellman equation from which we obtain G_{rmn} as a function of α_{rmn} , β_{rmn} , and a . The solution is best understood as a two-armed bandit (Gittins 1989).

Consider first an arm with known payoff probability, G_{rmn} . If we always select this arm, the expected reward in each and every period is G_{rmn} times the reward for success. Without loss of generality, normalize the reward for success to 1.0. If we discount future periods by a factor of a per period, the net present value is computed with the closed form of a geometric series: $G_{rmn} / (1 - a)$. The reward for selecting an uncertain arm is more complicated to derive because

³ Intuitively, we find an arm with certain expected payoffs such that we are indifferent between the uncertain arm and the certain arm. We then compare the corresponding certain arms and choose the arm with the highest payoff.

each success or failure updates our beliefs about the probability of success.

Following standard dynamic programming notation we let $R(\alpha_{r_{mn}}, \beta_{r_{mn}}, a)$ be the value of acting optimally. To act optimally, we must choose one of two actions, the known arm or the uncertain arm. When we select the uncertain arm, we get a success (with probability $\alpha_{r_{mn}}/(\alpha_{r_{mn}} + \beta_{r_{mn}})$) or a failure (with probability $\beta_{r_{mn}}/(\alpha_{r_{mn}} + \beta_{r_{mn}})$). If we observe a success, we get the payoff of 1.0 plus the discounted payoff we will receive for acting optimally in the future. The success also updates our beliefs about the future. Specifically, $\alpha_{r_{mn+1}} = \alpha_{r_{mn}} + 1$ and $\beta_{r_{mn+1}} = \beta_{r_{mn}}$. Thus, we expect a discounted reward of $1 + aR(\alpha_{r_{mn}} + 1, \beta_{r_{mn}}, a)$ when we observe a success. By similar reasoning, we expect a discounted reward of $aR(\alpha_{r_{mn}}, \beta_{r_{mn}} + 1, a)$ when we observe a failure. Putting these rewards together we calculate the expected reward of an uncertain arm as $\alpha_{r_{mn}}/(\alpha_{r_{mn}} + \beta_{r_{mn}})[1 + aR(\alpha_{r_{mn}} + 1, \beta_{r_{mn}}, a)] + \beta_{r_{mn}}/(\alpha_{r_{mn}} + \beta_{r_{mn}})aR(\alpha_{r_{mn}}, \beta_{r_{mn}} + 1, a)$. Our strategy is to choose the arm with the highest expected discounted profit; hence the Bellman equation becomes

$$R(\alpha_{r_{mn}}, \beta_{r_{mn}}, a) = \max \left\{ \frac{G_{r_{mn}}}{1-a}, \frac{\alpha_{r_{mn}}}{\alpha_{r_{mn}} + \beta_{r_{mn}}} [1 + aR(\alpha_{r_{mn}} + 1, \beta_{r_{mn}}, a)] + \frac{\beta_{r_{mn}}}{\alpha_{r_{mn}} + \beta_{r_{mn}}} aR(\alpha_{r_{mn}}, \beta_{r_{mn}} + 1, a) \right\}. \quad (1)$$

Equation (1) has no analytic solution, but we can readily compute Gittins indices with a simple iterative numeric algorithm.⁴ We illustrate $G_{r_{mn}}$ as a function of n in Appendix 3. As expected, the indices behave in an intuitive manner. If uncertainty is high (n small), exploration is valuable and $G_{r_{mn}}$ exceeds $\alpha_{r_{mn}}/(\alpha_{r_{mn}} + \beta_{r_{mn}})$ substantially. As we observe more website visitors, $G_{r_{mn}}$ decreases as a function of n . As $n \rightarrow \infty$, the expected rewards become known and $G_{r_{mn}}$ converges to $\alpha_{r_{mn}}/(\alpha_{r_{mn}} + \beta_{r_{mn}})$. The discount rate, a , is constant for our application, but if a were to increase, we would value the future more, and $G_{r_{mn}}$ would increase to make exploration more attractive.

Given a we precompute a table of indices for the values of $\alpha_{r_{mn}}$ and $\beta_{r_{mn}}$ that we expect to observe in the BT application, using interpolation if necessary. The $\alpha - \beta$ table is made manageable by recognizing that $G_{r_{mn}}$ converges to $\alpha_{r_{mn}}/(\alpha_{r_{mn}} + \beta_{r_{mn}})$ as the number of visitors gets large.

4.1. Is Gittins' Solution Reasonable for BT's Website?

It is not uncommon for a retail website to have 100,000 visitors per annum. With so many visitors it

is likely to be valuable to explore different morphs for early visitors so that BT can profit by providing the correct morph to later visitors. Suppose BT values future capital with a 10% discount per annum and suppose 100,000 visitors are spread evenly throughout the year. Then the effective discount from one visitor to the next is 1/100,000th of 10%, suggesting an implied discount factor of $a = 0.999999$. Even if visitors are spread among 16 cognitive-style segments, the effective discount factor is much closer to 1.0 than the discount factors used in typical Gittins applications (e.g., clinical trials, optimal experiments, job search, oil exploration, technology choice, and research and development; Jun 2004). With a so close to 1.0, we expect a Gittins strategy to entail a good deal of exploration. It is a valid fear that such exploration might lead to costly false morph assignments more so than a null strategy of one website for everyone. (The Gittins strategy is optimal if we allow morphing. The question here is whether morphing per se is reasonable in the face of issues outside our model. That is, is there a noticeable improvement relative to a no-morph strategy?⁵)

To address this practical implementation question, we use an a appropriate to BT's experimental website and we generate synthetic visitors who behave as we expect real visitors to behave. Our simulations are grounded empirically based on an experimental website. Full-scale implementation is planned, but production results are likely a year or more away.

We estimate real behavior by exposing a sample of 835 website visitors to one of eight randomly chosen morphs and observing their stated purchase probabilities. We measure cognitive styles with an intrusive question bank and estimate p_{rm} for each segment $\times$ morph combination. (Details are in §§7–9.) For example, to simulate one cognitive-style segment we used empirically derived probabilities {0.2996, 0.2945, 0.4023, 0.3901, 0.2624, 0.2606, 0.3658, 0.3580}; for morphs $m = 0$ to 7. For each synthetic visitor we generate a purchase using the probability that matches the morph assigned by the Gittins strategy. We generate 5,000 visitors in each of 16 cognitive-style segments (80,000 in total). This is well within the number of visitors to BT's website.

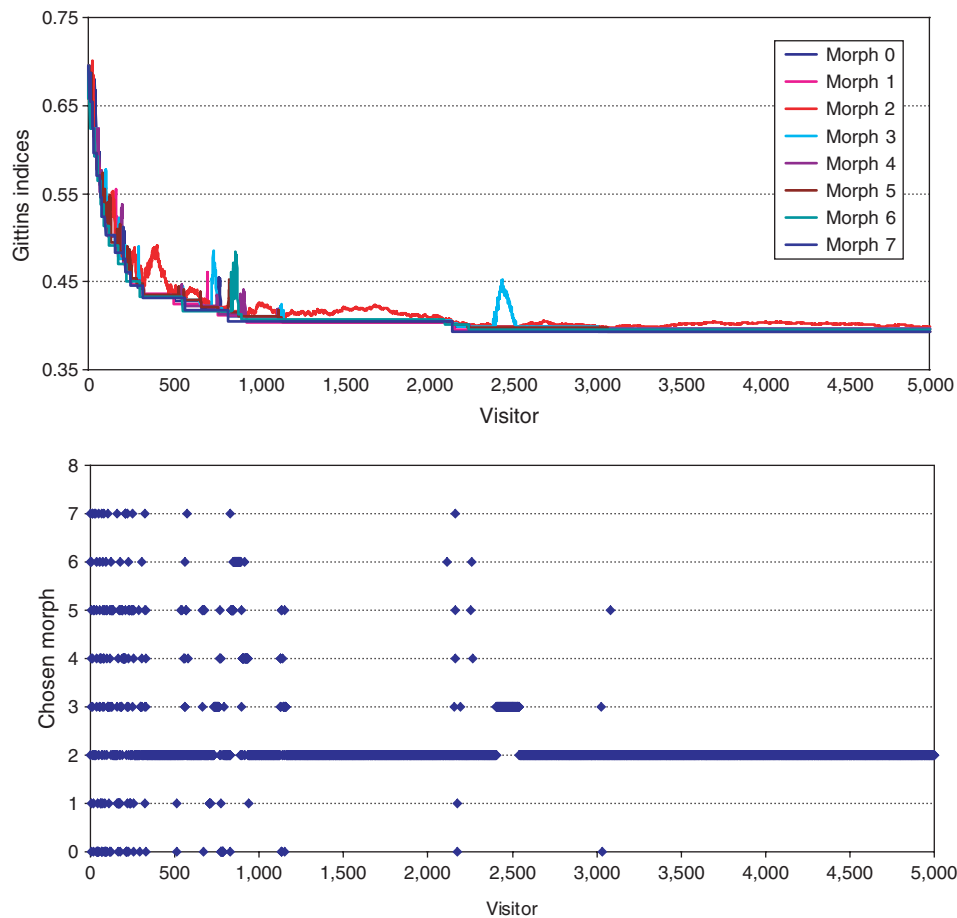
We seek a conservative test. As a lower bound, we start the system with equally likely prior probabilities that do not vary by morph and we begin with low precision beta priors. To avoid ties in the first morph assignment, we perturb the prior means randomly.

Figure 3 illustrates website morphing for a sample cognitive-style segment. The first panel plots the evolution of the Gittins indices; the second panel plots the

⁴ We are indebted to Professor John Gittins for sharing his code with us.

⁵ However, we would still have to be able to identify the no-morph strategy—itsself a Gittins problem.

Figure 3 Evolution of the Dynamic Programming Loop



morph chosen by the system. The Gittins indices for each of the eight morphs all start close to 0.7, which is significantly higher than the best-morph probability (approximately 0.4). The larger values of the indices reflect the option value of our uncertainty about the true probabilities. For the first few hundred visitors, the system experiments with various morphs before more or less settling on Morph 2 (red line). However, the system still experiments until about the 1,200th visitor. Around the 2,500th visitor the system flirts with Morph 3 (cyan line) before settling down again on Morph 2. This blip around the 2,500th visitor stems from random variation—a run of luck in which visitors purchased after seeing Morph 3. Morph 3’s probability of buying is 0.3901. It is close to, but not better than, Morph 2’s value of 0.4023. The system settles down after this run of luck, illustrating that the long-term behavior of the Gittins strategy is robust to such random perturbations.

Because the Gittins strategy is optimal in the presence of uncertainty, we can calculate the cost of uncertainty for this cognitive-style segment. The best morph for this segment is Morph 2 with an expected reward of 0.4023 times BT’s profit per sale. If we had perfect information, we would always

choose Morph 2 for this segment and achieve this expected reward. Because the Gittins strategy does not have perfect information, it explores other morphs before settling down on Morph 2. Despite the cost of exploration, the Gittins strategy achieves an expected reward of 0.3913, which is 97.2% of what we could have attained had perfect information been available. This is typical. When we average across cognitive-style segments we achieve an expected reward of 97.3% of that obtainable with perfect information.

We can also estimate the value of morphing. A website that is not designed with cognitive styles in mind is equivalent to one for which BT chooses one of the morphs randomly. In that case, the expected reward is 0.3292 times BT’s profit per sale. The Gittins strategy improves profits by 18.9%. Even if we had perfect information on purchase probabilities, we would only do 22.2% better. Strong priors (see §9) improve the Gittins strategy slightly—a 19.7% improvement relative to no morphing. These results illustrate the potential improvements that are possible by using the Gittins strategy to identify the best morph for a segment (assuming we knew to which segment the visitor belonged). We now extend our framework

to deal with uncertainty in cognitive-style-segment membership.

5. Dynamic Programming When Cognitive Styles Are Inferred (POMDP)

It is not feasible for BT to use an intrusive cognitive-style assessment on its production website. However, it is feasible to infer cognitive styles from visitors' clickstreams with the Bayesian inference loop. We demonstrate in §6 how the clickstream provides a posterior probability, $q_{rn} = f(r_n | \bar{y}_n, \Omega, \bar{c}_{kjn}s)$, that visitor n is in cognitive-style segment r_n . Because the state space of cognitive styles is only partially observable, the resulting optimization problem is a POMDP. The state space is Markov because the full history of the multiple-visitor process is summarized by r_n , the α_{rmn} s, and the β_{rmn} s. The POMDP cannot be solved optimally in real time, but good heuristics achieve near-optimal morph-assignment strategies. To incorporate uncertainty on cognitive styles, we make three modifications.

First, the Gittins strategy defines a unique morph per visitor and assumes the visitor makes a purchase decision after having experienced that morph. The outcome of the purchase-decision Bernoulli process is an independent random variable conditioned on the morph seen by a visitor. Although we do not know with certainty to which cognitive-style segment to assign this observation, we do know the probability, q_{rn} , that the observation, δ_{mn} , updates the r th cognitive-style segment's parameters.⁶ Because the beta and binomial distributions are conjugate, Bayes' theorem provides a means to use q_{rn} and δ_{mn} to update the beta distributions:

$$\begin{aligned}\alpha_{rmn} &= \alpha_{rm, n-1} + \delta_{mn} q_{rn}, \\ \beta_{rmn} &= \beta_{rm, n-1} + [1 - \delta_{mn}] q_{rn}.\end{aligned}\quad (2)$$

Second, following Krishnamurthy and Mickova (1999; hereafter referred to as KM) we compute an expected reward over the distribution of cognitive-style segments (the vector of probabilities q_{rn}) as well as over the posterior beta distribution with parameters α_{rmn} and β_{rmn} . KM demonstrate that while the full POMDP can be solved with a complex index strategy, a simple heuristic solution, called an Expected Gittins Index (EGI) strategy, achieves close to 99% of optimality. KM's EGI algorithm replaces G_{rmn} with EG_{mn} and chooses the morph with the largest EG_{mn} , where

$$EG_{mn} = \sum_{r=0}^{15} q_{rn} G_{rmn}(\alpha_{rmn}, \beta_{rmn}). \quad (3)$$

⁶ Because r_n is now partially observable, we have returned to the δ_{mn} notation, dropping the r subscript. To simplify exposition we continue to assume temporarily that the visitor experienced the m th morph for the entire visit.

For BT's experimental websites we cannot guarantee that KM's EGI solution will be within 99% of optimality (as in their problems). Instead, we bound the EGI's performance with comparisons to the expected rewards that would be obtained if BT were able to have perfect information on cognitive styles. The EGI solution does quite well (details are in §6).

Third, even if the website morphs once per visitor, the visitor sees the best initial morph, m_o , for part of the visit and the EGI-assigned POMDP morph, m^* , for the remainder of the visit. To update the EGI we must assign the visitor's purchase (or lack thereof) to a morph. The appropriate purchase-assignment rule is an empirical issue. If the number of clicks on m^* is sufficiently large relative to the number of clicks on m_o , then we assign the purchase to m^* and update only the indices for morph m^* . (We use the same rule if the last morph, m^* , has the strongest effect on purchase probabilities.) Alternatively, we can assign the purchase-or-not observation to m_o and m probabilistically based on the number of clicks on each morph. Other rules are possible. For example, we might weight later (or earlier) morphs more heavily or we might condition $p_{r\{m1, m2, m3, \dots\}}$ on a sequence of morphs, $\{m_1, m_2, m_3, \dots\}$. For our data we obtain good results by assigning the observation to m^* . Fortunately, for the BT experimental website, simulations with proportional purchase-assignment rules suggest that the performance of the system is robust with respect to such assignment rules.⁷ We leave further investigation of purchase-assignment rules to future research.

6. Inferring Cognitive Styles—A Bayesian Loop

BT's website is designed to provide information about and sell broadband service. Asking respondents to complete a lengthy questionnaire to identify their cognitive styles prior to exploring BT's website is onerous to visitors and might lower, rather than raise, sales of broadband service. Thus, rather than asking website visitors to directly describe their cognitive styles, the Bayesian loop infers cognitive styles. Specifically, after observing the clickstream, $\bar{y}_n$, and the click-alternative characteristics, $\bar{c}_{kjn}s$, we update the probabilities that the n th visitor belongs to each of the cognitive-style segments ($q_{rn}s$). (Although the $\bar{c}_{kjn}s$ depend on the initial morph, m_o , seen by the n th visitor, we continue

⁷ For example, with a last-morph assignment rule we obtain a mean posterior probability (q_{rn}) of 0.815 and a median posterior probability of 0.995. With a proportional-morph assignment rule, the mean is higher (0.877) but the median lower (0.970). The resulting rewards are quite close. To explore this issue empirically, we might seek data in which we assign both m_o and m^* randomly rather than endogeneously using the EGI solution to the POMDP.

to suppress the m_o subscript to keep the notation simple.)

We assume the n th visitor has unobserved preferences, $\tilde{u}_{k|n}$, for click-alternatives based on the click-alternative characteristics, $\tilde{c}_{k|n}$ s, and based on his or her preference weights, $\tilde{\omega}_{r_n}$, for those characteristics. We assume that preference weights vary by cognitive-style segment. (Recall that Ω is the matrix of the $\tilde{\omega}_{r_n}$ s. Temporarily assume it is known.) We express these unobserved preferences as $\tilde{u}_{k|n} = \tilde{c}_{k|n}'\tilde{\omega}_{r_n} + \tilde{\varepsilon}_{k|n}$, where $\tilde{\varepsilon}_{k|n}$ has an extreme-value distribution. Conditioned on a cognitive-style segment, r_n , the probability that we observe $\tilde{y}_{k|n}$ for the k th click by the n th visitor is

$$f(\tilde{y}_{k|n} | \tilde{c}_{k|n}, r_n, \Omega) = \prod_{j=1}^{J_k} \left(\frac{\exp[\tilde{c}_{k|n}'\tilde{\omega}_{r_n}]}{\sum_{\ell=1}^{J_k} \exp[\tilde{c}_{k|\ell n}'\tilde{\omega}_{r_n}]} \right)^{y_{k|n}}. \quad (4)$$

After we observe K_n clicks, the posterior distribution for cognitive-style segments is given by Bayes' theorem:

$$q_{rn} = f(r_n | \tilde{y}_n, \tilde{c}_{k|n}s, \Omega) \\ = \frac{\prod_{k=1}^{K_n} f(\tilde{y}_{k|n} | \tilde{c}_{k|n}, r_n, \Omega) q_o(r_n)}{\sum_{r=0}^{15} \prod_{k=1}^{K_n} f(\tilde{y}_{k|n} | \tilde{c}_{k|n}, r, \Omega) q_o(r)}, \quad (5)$$

where the $q_o(r_n)$ are the prior probabilities that the n th visitor belongs to cognitive-style segment r_n . Computing the q_{rn} s and the corresponding EG_{mn} s is sufficiently fast (~ 0.4 seconds; dual processor, 3 GHz, 4 GB RAM); visitors notice no delays on BT's experimental website.

Equations (4) and (5) require prior probabilities, $q_o(r)$, and estimates of the preference matrix, Ω . The click-alternative characteristics, $\tilde{c}_{k|n}$ s, are data. We obtain $q_o(r_n)$ and Ω from a priming study as described in §7. Because we use Bayesian methods to estimate Ω , it is theoretically consistent to update the q_{rn} s using the full posterior. Unfortunately, this is not yet practical because computation time is roughly linear in the number of samples from Ω 's posterior distribution. For example, with only 15 samples from the posterior it took 6.5 seconds to compute the EG_{mn} s—too long between clicks in a production setting. Furthermore, 15 samples is far too few to integrate effectively over the 50-element posterior distribution of Ω . This practical barrier might fall with faster computers and faster computational methods.⁸

In practice, if we identify new types of click-alternative characteristics or if BT feels that Ω has changed because of unobserved shocks, then selected

visitors can be invited to complete the priming-study questionnaire to provide data to update Ω .⁹ At any time, we can update $q_o(r_n)$ based on averaging the posterior q_{rn} over n .

Summary of the Gittins and Bayesian Loops

For each visitor, we update q_{rn} after each click. EG_{mn} predicts the best morph based on these q_{rn} s. After a set of initial clicks we morph the website to that best morph. After observing a purchase occasion we update the α_{rmn} s and β_{rmn} s for the next visitor. We use these updated α_{rmn} s and β_{rmn} s to update the Gittins indices and continue to the next visitor. As n gets sufficiently large, the system automatically learns the true p_{rm} s.

6.1. The Effect of Imperfect Cognitive-Style Identification

In §5 we found that the cost of uncertainty in segment $\times$ morph probabilities reduced the optimal solution to 97.2%, of that which we would obtain if we had (hypothetical) perfect information. The EGI solution to the POMDP should achieve close to the optimal morph assignment in the face of uncertainty on both segment $\times$ morph probabilities and cognitive styles, but that is an empirical question. To examine this question we compare the performance of the POMDP EGI solution to four benchmarks.¹⁰ Rewards are scaled such that 1.0000 means that every visitor purchases broadband service. The benchmarks are as follows:

- A website without the Gittins loop and no knowledge of cognitive styles.¹¹ The expected reward is 0.3205.
- A website with the Gittins loop, but no customization for cognitive-style segments. The expected reward is 0.3625.
- A website with the Gittins loop and (hypothetical) perfect information on cognitive-style segments. The expected reward is 0.3879.
- A website with (hypothetical) perfect knowledge of purchase probabilities and cognitive-style segments. The expected reward is 0.3984.

To compare the EGI solution to these benchmarks we begin with a scenario that illustrates the potential of the POMDP. We create synthetic Web pages

⁸ We tested a 15-sample strategy with synthetic data. The results were virtually indistinguishable from those we obtained using the posterior mean for Ω . Testing with large numbers of samples is not feasible at this time.

⁹ This last step adds no new conceptual challenges and incurs a modest, but not trivial, cost. BT has not yet seen a need to collect these additional data for its experimental website. The current implementation assumes that preferences vary by cognitive styles but are homogeneous within cognitive-style segment.

¹⁰ Figure 3 and the corresponding Gittins improvements in §4 are for a representative cognitive-style segment. The benchmarks cited here are based on the results of all 16 cognitive-style segments.

¹¹ Without information on cognitive styles or the Gittins loop, BT must select one of the eight morphs at random.

Table 1 Benchmark Comparison of the Inference and Optimization Systems

	Expected reward	Improvement (%)	Efficiency (%)	Relative efficiency (%)
No Gittins loop nor knowledge of cognitive styles.	0.3205	0.0	80.4	0.0
No morphing. Website chosen optimally by Gittins loop.	0.3625	13.1	91.0	53.9
Morphing: Match characteristics to cognitive-style segment.				
Bayesian inference of cognitive styles (10 clicks).	0.3844	19.9	96.5	82.0
Bayesian inference of cognitive styles (50 clicks).	0.3865	20.6	97.0	84.7
Perfect information on cognitive styles, Gittins loop.*	0.3879	21.0	97.4	85.5
Perfect information on style and purchase probabilities.*	0.3984	24.3	100	100

*Upper bounds. BT does not have perfect information on cognitive styles or purchase probabilities.

($\tilde{c}_{jkn}$ s) that provide clear choices in click-alternative characteristics both among and within morphs. In the simulations we know each customer's cognitive style, r . We create synthetic clickstreams from representative ω_r s by making multinomial draws from the random-utility model in Equation (4). After 10 clicks, we use the Bayesian loop to update q_{rm} and choose an optimal morph based on the EGIs. The synthetic customer then purchases a broadband service with probability p_{rm} , where r is the true cognitive state and m is the morph provided by the EGIs. (The EGIs may or may not have chosen the best morph for that synthetic customer.) Based on the observed purchase (δ_{mn}), we update the α_{rmn} s and β_{rmn} s and go to the next customer. We simulate 80,000 customers (5,000 customers per cognitive-style segment). As the number of clicks per customer increases, we expect the (Bayesian) posterior q_{rm} s to converge toward certainty and the rewards to converge toward those based on (hypothetical) perfect cognitive-style-segment information. Thus, for comparison, we include a 50-click simulation even though 50 clicks are more clicks than we observe for the average BT website visitor.

This simulation illustrates the potential of the EGI solution. It corresponds to a second generation website (Gen-2) that is now under development. The first-generation (Gen-1) BT experimental website was, to the best of our knowledge, the first attempt to design a website that morphs based on cognitive-style segments. We learned from our experience with that website. There were sufficient differences among morphs to identify p_{rm} easily with the Gittins loop; however, the relative similarity between click-alternatives within a morph meant that the Bayesian loop required more click observations than anticipated. We return to the Gen-1 website after we describe fully the empirical priming study (see §§7 and 8). The empirical insights obtained by comparing the Gen-1 and Gen-2 simulations are best understood based on the Ω estimated from the data in the priming study. (The Gen-1 Bayesian-loop improvements in revenue that we report in §10 are less dramatic but not insignificant from BT's perspective.)

In Table 1 we compare the Bayesian loop to the four benchmarks with three metrics. "Improvement" is the percent gain relative to the baseline of what would happen if a website were created without any attempt to take cognitive styles into consideration. The 10-click Bayesian/Gittins loop improves sales by 19.9%. "Efficiency" is the percentage of sales relative to that which could be obtained with perfect knowledge. The 10-click Bayesian/Gittins loop attains 96.5% of that benchmark. "Relative efficiency" is the percent gain relative to the difference in the lower and upper benchmarks. The 10-click Bayesian/Gittins loop attains an 82.0% relative efficiency.

Based on 10 clicks the Bayesian loop can identify most cognitive states. The median posterior probability (q_{rm}) is 0.898; the lower and upper quartiles are 0.684 and 0.979, respectively. However, on four of the cognitive states the Bayesian loop does not do as well; posterior probabilities are in the range of 0.387 to 0.593. If we were to allow more clicks (50 clicks) than we observe for the average website visitor, the posterior probabilities converge toward certainty. Based on 50 clicks the median and upper quartile are both 1.00, while the lower quartile is 0.959. The efficiency is 97.0%—very close to what BT would obtain if it had perfect information on cognitive styles (97.4%).

We estimate the marginal contribution of the Gen-2 Bayesian loop using revenue projections based on discussions with managers at the BT Group. (Gen-1 results are discussed in §10.) A 20% increase in sales corresponds to an approximately \$80 million increase in revenue. The Gittins loop projects a gain of approximately \$52.3 million by finding the best morph even without customization. The 10-click Bayesian loop adds another \$27.4 million by customizing the look and feel of the website based on posterior cognitive-style-segment probabilities. This is within \$2.6 million of what could be obtained with 50 clicks. Perfect information on cognitive-style segments would add yet another \$1.8 million, bringing us to \$84.1 million. These potential improvements are not insignificant to BT. However, we must caution the reader that BT has not yet implemented a Gen-2 website, and

the Gen-1 website is still experimental. Many practical implementation issues remain before these gains are achieved.

7. Data to Prime the Automated Inference Loops

We now describe the priming study for the experimental BT website. Although the morphing theory of §§2–6 can be applied to a wide range of websites, the priming study is an integral component of the BT application. It provides priors for the α_{rmoS} , β_{rmoS} , and $q_o(r_n)$ s and data with which to estimate preference weights (Ω) for website characteristics.

7.1. Priming Study—Questionnaires to Potential BT Website Visitors

Using a professional market research company (Applied Marketing Science, Inc.) and a respected British online panel (Research Now), we invited current and potential broadband users to complete an online questionnaire that combined BT's experimental website with a series of preference and cognitive-style questions. This sampling strategy attempts to obtain a representative sample of potential visitors to BT's broadband website. Because these data are used to calibrate key parts of the preference model, it is important that this sample be as representative as is feasible. Within a cognitive-style segment, we seek to assure that any response bias, if it exists, is not correlated with $\vec{\omega}_{r_n}$. Fortunately, with sufficient production-website data, the Gittins and Bayesian loops should self-correct for response biases, if any, in segment $\times$ morph probabilities and/or cognitive-style segment-membership priors.

A total of 835 respondents completed the questionnaire. Because the questionnaire was comprehensive and time consuming, respondents received an incentive of £15. The questionnaire contained the following sequential sections:

- Respondents answer questions to identify whether they are in the target market.
- Respondents identify which of 16 broadband providers they would consider and provide initial purchase-intention probabilities for considered providers.
- Respondents are given a chance to explore one of eight potential morphs for the BT website. The morphs were assigned randomly, and respondents were encouraged to spend at least five minutes on BT's experimental website.
- Respondents provide post-visit consideration and purchase-intention probabilities.¹²

- Respondents are shown eight pairs of websites that vary on three basic characteristics. They are asked to express their preferences between the pairs of websites with a choice-based conjoint analysis-like exercise. These data augment clickstream data when estimating Ω .

- Respondents complete established scales that the academic literature suggests measure cognitive styles. The questionnaire closes with demographic information.

Reaction to the experimental BT websites was positive. Respondents found the websites to be helpful, accurate, relevant, easy to use, enjoyable, and informative (average scores ranging from 3.2 to 3.8 out of 5.0). On average, respondents clicked more than 10 times while exploring the websites, with 10% of the respondents clicking over 30 times.

7.2. Cognitive Style Measures

Figure 4 provides 10 of the 13 scales that we used to measure cognitive styles. We chose these scales based on prior literature as the most likely to affect respondents' preferences for website characteristics. We expect these scales to be a good start for website applications. To encourage further development, the Technical Appendix, available at <http://mktsci.pubs.informs.org>, provides a taxonomy of potential cognitive styles.

We expected these scales to identify whether the respondent was analytic or holistic, impulsive or deliberative, visual or verbal, and a leader or a follower. The analytic versus holistic dimension is widely studied in psychology and viewed as being a major differentiator of how individuals organize and process information (Riding and Rayner 1998, Allinson and Hayes 1996, Kirton 1987, Riding and Cheema 1991). Researchers in both psychology and marketing suggest that cognitive styles can be further differentiated as either impulsive or deliberative (Kopfsstein 1973, Siegelman 1969). With a slight rescaling three cognitive reflection scales developed by Frederick (2005) differentiate respondents on the impulsive versus deliberative dimension.¹³ Other scales measure visual versus verbal styles, a key cognitive concept in psychology (Harvey et al. 1961, Paivio 1971, Riding and Taylor 1976, Riding and Calvey 1981). This dimension is particularly relevant to website design where the trade-off between pictures and text is an important design element. Although leadership is not commonly a cognitive-

preferences, Ω , should not be affected by any induced demand artifacts. Any demand artifacts affect primarily the priors. Fortunately, the Gittins loop is relatively insensitive to prior probabilities.

¹³ For example, "A bat and a ball cost \$1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost?" The impulsive answer is 10¢; all other answers are considered to be deliberative.

¹² Because respondents see only the BT website, we attempt to minimize demand artifacts by renormalizing the data. Click-characteristic

Figure 4 Example Measures of Cognitive Styles

Broadband Usage and Attitude

Now we would like to learn a little bit more about you in order to understand your personal decision style when choosing between options. Please answer to what extent you agree or disagree with the statements below to best represents how you think or feel.

For each statement below, please indicate the number that best indicates your level of agreement using the 5-point scale below.
[CHOOSE ONE ONLY FOR EACH STATEMENT]

	Strongly Disagree			Strongly Agree	
	1	2	3	4	5
I prefer to read text rather than to listen to a lecture	☐	☐	☐	☐	☐
I enjoy deciphering graphs, charts, and diagrams	☐	☐	☐	☐	☐
I will read an explanation of a graph/chart before I try to understand the graph/chart on my own	☐	☐	☐	☐	☐
I see what I read in mental pictures	☐	☐	☐	☐	☐
I am detail oriented, and start with the details in order to build a complete picture	☐	☐	☐	☐	☐
I tend to see problems in their entirety and start by integrating pieces from different areas	☐	☐	☐	☐	☐
I find it is easy to make decisions for others and to command and direct others to take certain actions	☐	☐	☐	☐	☐
In a group conversation, I usually speak the most	☐	☐	☐	☐	☐
I have held a great deal of leadership positions in my life	☐	☐	☐	☐	☐
My confidence level is higher than most other people's	☐	☐	☐	☐	☐

NEXT

style dimension in psychology, we included leadership scales because thought leadership has proven important in the adoption of new products and new information sources (Rogers 1962, Rogers and Stanfield 1968, von Hippel 1988). To the extent that we included scales that do not distinguish cognitive styles, our empirical analyses will find null effects. Additional scales can be explored in future research. Our results are a conservative indicator of what is feasible with improved scales.

Although the scales are well established in the literature, we began with construct tests using our data. We used exploratory factor analysis and confirmatory reliability analyses to reduce the 13 scales (10 scales from Figure 4 plus the 3 impulsive versus deliberate scales) to four cognitive dimensions. (See Braun et al. 2008 for greater detail on scale development and analysis.) For the BT data, impulsive versus deliberative and leader versus follower were measured with sufficient reliability (0.55 and 0.80, respectively); analytic versus holistic and visual versus verbal combined to a single construct (0.56 reliability). The analyses identified a fourth dimension: a single scale, reader versus

listener. We suspect that this reader versus listener scale was driven by the nature of the broadband service websites that often give respondents a choice of reading text or tables or listening to an advisor. Although multi-item scales are more common in the literature, recent research recognizes the corresponding advantages of single-item scales (Bergkvist and Rossiter 2007, Drolet and Morrison 2001). Based on this research we include this single-item scale as a fourth cognitive-style dimension.

Although some of these reliabilities are lower than we would like, this reflects the challenges in measuring cognitive styles and, for our analytic models, adds noise to the estimation of Ω and to the Bayesian loop. Fortunately, the constructs as measured appear to affect purchase probabilities (see Braun et al. 2008). In summary, we identified four empirical constructs to measure respondents' cognitive styles:

- leader versus follower,
- analytic/visual versus holistic/verbal,
- impulsive versus deliberative,
- (active) reader versus (passive) listener.

Using median splits, we define $16 = 2 \times 2 \times 2 \times 2$ cognitive-style segments.¹⁴

7.3. Click-Alternative Characteristics

There are four sources of variation in click-alternative characteristics. First, the morphs themselves vary on three basic dimensions. Second, click-alternatives within the morphs vary on the same three dimensions. Third, there are functional characteristics of click-alternatives, for example, whether a link provides general information (of potential interest to holistic respondents). Fourth, the home page of the experimental BT website gives the respondent a choice of four content areas. We expect visitors with different cognitive styles to vary in their desire to visit different content areas on their first click.

7.3.1. Basic Characteristics of a Morph. Based on the literature cited above we chose three basic click-alternative characteristics that were likely to distinguish morphs and click-alternatives within morphs. These characteristics were used to design the basic structures (backbones) of the BT experimental websites based on initial hypotheses about the variation among cognitive-style segments in preferences for characteristics. The characteristics varied on the following:

- graphical versus verbal (e.g., graphs and pictures versus text and audio),
- small-load versus large-load (e.g., the amount of information presented),
- focused content versus general content (e.g., a few recommended plans versus all plans).

The characteristics of the websites (morphs) that were shown (randomly) to each respondent at the beginning of the questionnaire and the characteristics of the pairs of websites shown in the choice-based conjoint-like exercise were designed to be distinguished on these basic click-alternative dimensions. Hence, we describe each morph by one of eight binary vectors, from $\{0, 0, 0\}$ to $\{1, 1, 1\}$. For example, the $\{1, 1, 1\}$ morph is graphic, focused, and small load. This binary notation is chosen to be consistent with the earlier notation of $m = 0$ to 7; e.g., $m = 0 \Leftrightarrow \{0, 0, 0\}$.

We invested considerable effort to design morphs that would match cognitive styles, and to some extent, we succeeded. One advantage of the EGI optimization is that asymptotically it will identify automatically the best morph for a cognitive-style segment even if that morph is not the morph that we expect to be best a priori. The system in Figure 2 is robust with respect to errors in website design. In fact, a serendipitous outcome of the priming study was a better

understanding of website design and the need for a Gen-2 experimental website.

7.3.2. Characteristics of Click Alternatives Within a Morph. We used five independent judges to rate the basic characteristics of each click-alternative, a methodology that is common in marketing (e.g., Hughes and Garrett 1990, Perreault and Leigh 1989, Wright 1973). The judges were trained in the task but otherwise blind to any hypotheses. The average reliability of these ratings was 0.66 using a robust measure of reliability (proportional reduction in loss; Rust and Cooil 1994). Like cognitive styles, click-alternative characteristics are somewhat noisy but should provide sufficient information for the Bayesian loop and the estimation of preference weights (Ω).

7.3.3. Functional Characteristics of Click Alternatives. We identified four functional characteristics that were likely to appeal differentially to respondents with different cognitive styles. These functional characteristics were represented with the following binary variables:¹⁵

- general information about BT (e.g., likely to appeal to holistic visitors),
- analytic tool that allows visitors to manipulate information (e.g., likely to appeal to analytic visitors),
- link to read a posting by another consumer (e.g., likely to appeal to followers),
- link to post a comment (e.g., likely to appeal to deliberative visitors).

7.3.4. Content Areas. The home page of the experimental BT website offered the visitor four content areas (advisor, community, comparisons, and learning center), each of which could be morphed. Figure 5 illustrates these four content areas. To test whether the content areas would appeal differentially to respondents based on their cognitive-style segments, we coded the content areas as binary variables. (We have three, rather than four, independent dummy variables for the four content areas.)

Together, the three types of click-alternative variations give us ten (10) click-alternative characteristics: three basic dimensions, four functional characteristics, and three of four content areas.

8. Estimation of Click-Alternative Preferences, Ω , from the Priming Data

The Bayesian inference loop uses visitors' click-streams to compute posterior probabilities for

¹⁴ The Gittins inference/optimization loop is based on discretely many cognitive-style segments (r_n). Future research might explore more continuous cognitive-style descriptions of website visitors.

¹⁵ The BT experimental website also contained audio links, column headings, and a review of past information; however, these were collinear with the four primary characteristics. Generation 2 websites will be designed to make these and other characteristics as orthogonal as feasible given BT's primary goal of selling broadband service.

Figure 5 Broadband Advice Centre Home Page

BT Homepage | About BT

Broadband Advice Centre

Advocacy
COMPARE PLANS
ADVICE CENTRE
COMMUNITY
LEARNING CENTRE

Resources
BUILD PLAN
BID PRICE
BROADBAND NEWS
BROADBAND RATINGS
SPECIAL OFFERS
ALTERNATIVES
ABOUT THIS SITE
BUY NOW!

★ My History [Open](#)

Compare Plans

Provider	Price	Speeds	Reliability	Type
BT	£ 15	256 kb	Great	DSL
Wanadoo	£ 15	512 kb	Average	Satellite
Demon	£ 17	256 kb	Great	DSL
NetService	£ 22	256 kb	Great	Cable
AOL G	£ 24	256 kb	Good	Cable
Elite UK	£ 24	256 kb	Average	DSL
Atlas	£ 25	512 kb	Great	DSL
AOL P	£ 29	512 kb	Good	Cable

[More >](#)

Choose a range of providers [Go ▶](#)

Broadband Advisor

Broadband Advisor

Get an assistant to help you with your broadband needs
[More >](#)

[Get Advice on Broadband](#) [Go ▶](#)

Broadband Community

Right Now.

[Share and Consult views](#) [Go ▶](#)

Learning Centre

[Learn about Broadband](#) [Go ▶](#)

cognitive-style segments r_n . The posterior probabilities (q_{rn} , Equation (5)) require preference weights, Ω , for the click-alternative characteristics ($\tilde{c}_{k|n}$ s). We now address how we obtain from the priming data a posterior distribution for Ω . We can infer a posterior distribution for Ω because, in the priming data, we observe the respondent's cognitive-style segment directly. The inference problem is to infer Ω from $\{\tilde{y}_n s, \tilde{c}_{k|n} s, r_n s\}$.

We have two sources of data within the priming study. First, we observe each respondent's click-stream. Second, we augment each respondent's click-stream data with conjoint analysis-like data in which the respondent provides paired-comparison judgments for eight pairs of website pages. Because the latter choices among pairs of websites may not be derived from the same "utility" scale as choices from among click-alternatives, we allow for scale differences. Before we write out the likelihoods for each of the two types of data, we need additional notation.

8.1. Cognitive-Style-Segment Vector Notation

In §§2–6 we defined r_n as a scalar. This is a general formulation for the Gittins loop. It allows each cognitive-style segment to be independent of every other segment. In the BT application there are $2^4 = 16$ cognitive-style segments based on four binary cognitive-style dimensions. To reflect this interdependence among segments, we rewrite r_n as a 5×1 binary vector, $\tilde{r}_n$, where the first element is always equal to 1 and represents the characteristic-specific mean. Each subsequent element of $\tilde{r}_n$ reflects a deviation from that mean based on a cognitive-style dimension of the segment. For example, a member of cognitive-style segment $r_n = 0 \Leftrightarrow \tilde{r}_n = \{1, -1, -1, -1, -1\}$ is a follower, holistic/verbal, deliberative, and a listener; $r_n = 15 \Leftrightarrow \tilde{r}_n = \{1, 1, 1, 1, 1\}$ is a leader, analytic/visual, impulsive, and a reader. With this notation, we write characteristic preferences compactly as $\vec{\omega}_{r_n} = \Omega \tilde{r}_n$.

8.2. Clickstream Likelihood

Using the vector notation combined with the notation of §§2–6, the clickstream likelihood (CSL) is based on

Table 2 Results of Bayesian Updating on Website-Characteristic Preferences ($\gamma^{-1} = 0.17$)

	Mean effect	Leader versus follower	Analytic/visual versus holistic/verbal	Impulsive versus deliberative	Reader versus listener
Basic dimensions					
Graphical versus verbal	1.82	0.02	0.03	−0.01	−0.10
Small versus large load	− 1.85	0.07	−0.11	0.15	−0.02
Focused versus general	−0.09	−0.09	− 0.86	−0.04	0.27
Functional characteristics					
General information	0.08	−0.08	− 0.38	−0.07	0.10
Analytic tool	1.07	−0.07	0.02	−0.06	−0.03
Read a post	3.40	−0.17	0.05	0.08	−0.07
Post a comment	0.52	−0.02	0.13	−0.04	−0.13
Website areas					
Compare plans	2.56	−0.14	0.67	−0.02	−0.15
Virtual advisor	1.61	−0.12	0.27	−0.13	−0.06
Community	—	—	—	—	—
Learning center	0.13	−0.27	0.08	−0.04	0.11

Equation (5), except that Ω is unknown and the $\vec{r}_n$ are data. This likelihood assumes the unobserved errors are independent across clicks:

$$\text{CSL} = \prod_{n=1}^{835} \prod_{k=1}^{K_n} \prod_{j=1}^{J_k} \left(\frac{\exp[\vec{c}_{kjn}' \Omega \vec{r}_n]}{\sum_{\ell=1}^{J_k} \exp[\vec{c}_{k\ell n}' \Omega \vec{r}_n]} \right)^{y_{kjn}}. \quad (6)$$

8.3. Paired-Comparison Likelihood

Each respondent is presented with eight pairs of website pages that vary on the three basic morph characteristics of graphic versus verbal, focused versus general, and small versus large load. The eight pairs are chosen randomly from a 2^3 experimental design such that no pair is repeated for a respondent and left and right presentations were rotated randomly. The overall D-efficiency of this design is close to 100%. For each respondent, n , let $\vec{d}_{t1n}$ and $\vec{d}_{t2n}$ be the descriptions of the left and right website pages, respectively, for the t th pair on the three dimensions, and let s_{tn} indicate the selection of the left website page, $t = 1$ to 8. The respondent's preference for the left website page is based on the characteristics of the website pages. If $\tilde{\xi}_{tn}$ is an extreme-value measurement error, then the respondent's unobserved preference for the left website page is given by $\gamma(\vec{d}_{t1n}' - \vec{d}_{t2n}') \Omega \vec{r}_n + \tilde{\xi}_{tn}$. Note that we allow a differential scale factor, γ , to reflect possible differences between the clickstream and paired-comparison choice tasks. With this formulation, the paired-comparison likelihood (PCL) becomes the standard choice-based conjoint likelihood, which assumes that the unobserved errors are independent across paired-comparison choices:

$$\text{PCL} = \prod_{n=1}^{835} \prod_{t=1}^8 \left(\frac{\exp[\gamma \vec{d}_{t1n}' \Omega \vec{r}_n]}{\exp[\gamma \vec{d}_{t1n}' \Omega \vec{r}_n] + \exp[\gamma \vec{d}_{t2n}' \Omega \vec{r}_n]} \right)^{s_{tn}}. \quad (7)$$

Finally, we use the method of Train (2003) to match the variances in Equations (6) and (7) and to assure that Ω is scaled properly for both likelihoods.¹⁶

8.4. Posterior Distribution for Cognitive-Style Preferences

We combine Equations (6) and (7) with weakly informative priors, $g(\Omega, \gamma)$, on the unknown parameters to obtain a posterior distribution for the cognitive-style preferences and the scaling parameter. Equation (8) assumes that the unobserved errors in the clickstream are independent of the measurement errors in the paired comparison choices:

$$f(\Omega, \gamma \mid \vec{c}_{kjm}, \vec{d}_{t1n}, \vec{d}_{t2n}, s_{tn}, \vec{y}_n, \vec{r}_n \forall k, j, t, n) \\ \propto \text{PCL} * \text{CSL} * g(\Omega, \gamma). \quad (8)$$

From the 835 respondents in the priming study we observe 4,019 relevant clickstream choices and 6,680 paired-comparison choices. Samples from the posterior distribution of Ω and γ were generated using WinBUGS.¹⁷ Table 2 provides the posterior means of Ω . Appendix 2 provides the intervals between the 0.05 and 0.95 quantiles for the posterior distribution. Using the mean posterior probabilities alone, we explain 60.3% of uncertainty in the clickstream choices ($U^2 = 0.603$; Hauser 1978).

¹⁶ The standard deviations of the error terms, ε_{kjm} and ξ_{tn} , for the logit likelihoods determine the scale or “accuracy” of the parameter estimates. By allowing $\gamma \neq 1$, we automatically allow different standard deviations for the errors. Independence assumes the conjoint design is not endogenous (Hauser and Toubia 2005).

¹⁷ WinBUGS code and convergence details are available from the authors. As a check on the WinBUGS code, we also estimated Ω using classical methods (maximum likelihood estimation (MLE)). The Bayesian and MLE estimates were statistically equivalent.

We have highlighted in bold those coefficients for which the 0.05 to 0.95 quantile of the posterior distribution is either all positive or all negative. The lack of “significance” for the remaining coefficients might reflect insufficient variation in functional characteristics, the relative sparseness of data for the website areas (first click only), or unobserved variation.^{18, 19} We expect improved discrimination on BT’s Gen-2 websites. By creating more distinct click-alternative choices, the Gen-2 website will be better able to identify cognitive styles with only a few clicks.

On average, graphical content increases preference but small loads and focused content decrease preference. Analytic tools, consumer posts, plan comparisons, and virtual advisors are popular click choices by respondents. Respondents prefer to go first to website areas that compare plans and provide virtual advisors. There are also cognitive-style-specific effects: respondents who are holistic/verbal or readers prefer focused content. Although not quite “significant,” impulsive respondents prefer small information loads. The tendency to go first to plan comparisons and virtual advisors while avoiding general information appears to be a trait that distinguishes analytic/visual from holistic/verbal respondents.

In the spirit of Bayesian inference, we cautiously examine characteristics for which 80% of the posterior is either all positive or negative. In this case we would find that followers such as learning communities and listeners like to post comments and compare plans. Listeners also prefer verbal and general content and analytic/visual respondents prefer large information loads. We interpret these results, based on the Gen-1 experimental website, as hypotheses to be tested with Gen-2 websites and the corresponding priming studies.

9. Strong Priors for Gittins and Bayesian Loops

The priming study was based on a representative sample of potential visitors to BT’s experimental Gen-1 website. We can use these data to obtain strong priors with which to improve the performances of the

Gittins and Bayesian loops. For example, although the Gittins loop works well with equally likely priors on the beta parameters, the analyses of §4 suggest that we can achieve a slight improvement with stronger priors.

9.1. Prior Cognitive-Style-Segment Probabilities for the Bayesian Loop

Using the established scales we observed the cognitive-style segment, r_n , for every respondent in the representative sample. The empirical distribution of cognitive-style segments provides priors, $q_o(r_n)$, for the Bayesian loop.

9.2. Prior Purchase Probabilities for the Gittins Loop

In the priming study we observe directly each respondent’s purchase intentions. Thus, because we assigned each respondent randomly to one of the eight morphs and we inferred that respondent’s cognitive-style segment from the established scales, we have a direct estimate of the prior purchase probabilities for each segment $\times$ morph combination, $\bar{p}_{rmo}$. These direct estimates provide information on the prior beta parameters via $\bar{p}_{rmo} = \alpha_{rmo} / (\alpha_{rmo} + \beta_{rmo})$.

For the Gittins loop, we want the data to overwhelm the prior so we select a relatively small effective sample size, N_{rmo} , for the beta prior. Because $N_{rmo} = \alpha_{rmo} + \beta_{rmo}$ and because the variance of the beta distribution is

$$\alpha_{rmo}\beta_{rmo}/[(\alpha_{rmo} + \beta_{rmo})^2(\alpha_{rmo} + \beta_{rmo} + 1)],$$

we choose an approximate N_{rmo} by managerial judgment informed by matching the variance of the beta distribution to the variance of the observed purchase-intention probabilities. For our data we select $N_{rmo} \cong 12$.

9.3. Caveats and Practical Considerations

With sufficiently many website visitors from whom to observe actual purchase decisions, the $\bar{p}_{rmo}$ will converge to their true values and the priors will have negligible influence. Nonetheless, we sought to use the data more efficiently for obtaining strong priors for the Gittins and Bayesian loops. Our first practical consideration was sample size. With 835 respondents for 16 cognitive-style segments and eight morphs, the average sample size is small for each segment $\times$ morph estimate of $\bar{p}_{rmo}$. To make more efficient use of the data and smooth these estimates over the $r \times m$ cells, we used logistic regression. The explanatory variables were the basic characteristics of the morphs, the cognitive-style dimensions of the segments, and characteristic-dimension matches (e.g., small information loads for impulsive segments). The variance

¹⁸ We use the classical term “significance” as shorthand for the quantiles being either all positive or negative. We do this for ease of exposition recognizing the more subtle Bayesian interpretation.

¹⁹ Preferences vary across cognitive-style segments and the model does explain over 60% of the variation in clickstream choices. Future research might test more complex specifications subject to the need to update q_{rn} in real time. For example, if we specified a normal hyperdistribution over the 50 parameters in Table 2, updating q_{rn} would require extensive numerical integration (or simulated draws) in real time (e.g., 50 parameters $\times$ 16 segments $\times$ 10 clicks $\times$ 10 alternatives per click).

of $\bar{p}_{rmo}$ is also based on the smoothed estimates. See Braun et al. (2008) for further analyses.

Our second practical consideration in the priming study was the use of purchase intentions rather than observed purchases. In a production, website visitors self-select to come to BT's website; we expect such visitors are closer in time to purchasing broadband service than those recruited for the priming study. Although we were careful in recruiting to obtain a representative sample, we measured purchase intentions rather than observed purchases.²⁰ Purchase intentions have the benefit of obtaining a more discriminating measure from each respondent than 0 versus 1 purchase. However, purchase intentions are often subject to demand artifacts (e.g., Morwitz et al. 1993). For example, for nonfrequently purchased items, true probabilities tend to be linear in purchase intentions (Jamieson and Bass 1989, Kalwani and Silk 1982, Morrison 1979). To reduce the impact of potential scale factors, we normalized purchase intention measures relative to other broadband services and we used baseline benchmarks in Table 1 as quasi controls. Revenue increases are based on the relative efficiencies of the Gittins and Bayesian loops. Finally, because morphs were assigned randomly and each respondent saw only one morph, the relative differences between morphs are less sensitive to any demand artifacts.

10. Improvements and Further Applications

The development and testing of morphing websites is ongoing. BT is optimistic based on the Gen-1 priming study. Viewed as a feasibility test, the Gen-1 test identified a few website characteristics that could be matched to cognitive-style segments. The Gen-1 test also confirmed that website characteristics can affect purchase probabilities.

Before collecting data we did not know which of the eight morphs would maximize revenue. However, the Gittins loop alone (without morphing) identified the best website characteristics, implying an increase in revenue of \$52.3 million (Table 1 and §6). Section 6 also suggests that a Gen-2 website (designed to distinguish among cognitive styles clearly after 10 clicks) could increase revenues an additional \$27.4 million. Based on this "proof of concept," BT plans to implement the customer advocacy backbone, illustrated in Figures 1 and 5, and add Gen-2 morphing to the site as soon as feasible.

²⁰ As is appropriate ethically and legally, respondents were recruited with promises that we would not attempt to sell them anything in the guise of market research. Because of these guidelines we could not offer respondents the ability to sign up for a BT broadband plan.

In addition, Suruga Bank in Japan is developing and testing a morphing website to sell personal loans. The website morphs based on cognitive styles and cultural preferences such as hierarchical versus egalitarian, individual versus collective, and emotional versus neutral (Hofstede 1983, 1984, 2001; Trompenaars and Hampden-Turner 1997; Steenkamp et al. 1998).

10.1. Gen-1 Compared to Gen-2 Experimental Websites

The eight morphs in the Gen-1 experimental website were sufficiently varied in the way they affected purchase probabilities. However, the website characteristics within a morph (from which we identify cognitive-style segments) were not sufficiently varied in Gen-1. For example, the website areas on the Gen-1 home page were effective at distinguishing analytic/visual from holistic/verbal respondents (see Ω in Table 2), but less so on the other cognitive-style dimensions. The simulations in Table 1 assumed that website characteristics within a morph were more distinct leading to larger posterior means (Gen-2 Ω). (BT feels that such a website is feasible.)

To motivate Gen-2 development and to assess the Bayesian-loop gains for Gen-1, we resimulated the Bayesian loop with the Gen-1 Ω . (The Gittins-only-loop results remain unchanged.) With 10 clicks, 80,000 visitors, and a Gen-1 Ω , the expected reward is 0.3646. While the implied revenue increase is not insignificant for BT, the Gen-1 gains (total Gittins + Bayesian gains = \$54.9 million) are much smaller than the potential gains with a Gen-2 website (total gains = \$79.7 million). Interestingly, even the Gen-1 website could get substantially more revenue if it had infinitely many visitors such that the system learned almost perfectly the segment $\times$ morph purchase probabilities (p_{rm}). Gen-1 ($n = \infty$) could achieve \$75.7 million in additional revenues, close to that which Gen-2 achieves with 80,000 visitors.

11. Future Research to Improve the Theory and Practice of Morphing

Prior research and industry practice have demonstrated the power of self-selected branching, recommendations, and customized content (Ansari and Mela 2003, Montgomery et al. 2004). In this paper we explore the next step, changing the presentation of information to match each customer's cognitive style. The EGI solution to the POMDP enables us to explore different assignments of morphs to cognitive-style segments. The Bayesian updating enables customers to reveal their cognitive styles through their clickstreams. Together, the Gittins and Bayesian loops automate morphing (after a priming study).

Feasibility considerations required empirical trade-offs. We used segments of cognitive styles rather than continuously defined cognitive styles because the dynamic program requires finitely many “arms.” We morphed once per visit, in part, because we observe a single subscription decision per customer. We estimated homogeneous click-characteristic preference weights so that we could identify cognitive-style segments in real time. We used the posterior mean of Ω rather than sampling from the posterior distribution of Ω because we need to compute the EGI between clicks. Moreover, the priming study was based on a Gen-1 implementation. Each of these issues can be addressed in future applications.

BT was most interested in broadband subscriptions. In other applications, purchase amounts might be important. If purchase amounts are normal random variables, we can use normal priors rather than beta priors. Gittins (1979, pp. 160–161) demonstrates that this normal-normal case is also solved with an index strategy and provides algorithms to compute the normal-normal indices. Vermorel and Mohri (2005) explore a series of heuristic algorithms that perform well in online contexts. We easily extend the theory to a situation where we observe (1) whether a purchase is made and (2) the amount of that purchase. In this case we observe the normally distributed outcome conditioned on a Bernoulli outcome. This is a special case of “bandit-branching” as introduced by Weber (1992) and studied by Bertsimas and Niño-Mora (1996) and Tsitsiklis (1994). Using a “fair charge” argument, Weber shows that the value of a bandit-branching process can be computed by replacing the reward to a branch with its Gittins index. The index of a sales-then-sales-amount process becomes the product of the beta-Bernoulli and the normal-normal indices. All other considerations in Figure 2 remain the same. Recent developments in the bandit literature now make it feasible to include switching costs via fast generalized index heuristics (e.g., Dushonchet and Hongler 2006, Jun 2004).

Our application focused on cognitive styles. The literatures in psychology and learning posit that cognitive styles are enduring characteristics of human beings. If our EGI algorithm was extended to other marketing-mix elements besides website design, we might consider latent states that evolved randomly or based on marketing-mix elements. (See review in §3.) There are exciting opportunities to combine the advantages of HMMs or latent-class analysis with the exploration-exploitation trade-offs made possible with expected Gittins indices.

Acknowledgments

This research was supported by the MIT Sloan School of Management, the Center for Digital Business at MIT

(ebusiness.mit.edu), and the BT Group. The authors gratefully acknowledge the contributions of their industrial collaborators, research assistants, and faculty colleagues: Cinda Amyx, Theta Aye, Fabio Colacchio, Shane Frederick, Shirley Fung, Tilli Kalisky, Steven Kao, Robert Klein, Mangesh Kulkarni, Shelly Lau, Matthew Lawson, Antonio Lorenzon, Clarence Lee, Aman Narang, Gavin Patterson, Harold Salomann, Steven Stokols, Fareena Sultan, and Steven Whittaker. The authors thank the editor, area editor, and reviewers at *Marketing Science* for constructive comments that improved their paper.

Appendix 1. Notation

- a = amount by which future visitors are valued, the discount rate in the dynamic program;
- $\vec{c}_{k|n}$ = characteristics of the j th click-alternative of the k th click decision by visitor n ;
- $\vec{d}_{in1}, \vec{d}_{in2}$ = first three elements of $\vec{c}_{k|n}$; notation used for paired-comparison selections;
- EG_{mn} = expected Gittins index for the m th morph for visitor n ;
- $f(\cdot)$ = probability density function, usually the posterior distribution;
- $g(\cdot)$ = probability density function, usually a prior;
- G_{rmn} = Gittins index for the r th cognitive-style segment and the m th morph for visitor n ;
- j indexes click-alternatives;
- J_{kn} = number of click-alternatives at the k th click by visitor n ;
- k indexes clicks;
- K_n = number of clicks made by visitor n ;
- ℓ = used as an index in Equation (3); summation in the denominator;
- m = indexes morphs, $m = 1$ to 7 or, equivalently, m implies a binary representation;
- m_o = initial morph seen by website visitors;
- m_r^* = optimal morph for cognitive-style segment r ;
- n indexes visitors, used for both production visitors and priming-study respondents;
- N_{rm} = total number of visitors who see the m th morph and are in the r th cognitive-style segment;
- o indexes prior values, e.g., for $\alpha_{rm}, \beta_{rm}, p_{rm}, N_{rm}, m_o$;
- p_{rm} = probability that visitor n in the cognitive-style segment, r , will subscribe to BT when shown morph m ; $\bar{p}_{rmn}$ is the mean of the posterior for p_{rm} after the n th visitor; $\bar{p}_{rmo}$ is the mean of the prior for p_{rm} ;
- $\vec{p}$ = matrix of the p_{rm} s;
- $q_{rn} = f(r_n | \vec{y}_n, \vec{c}_{k|n}, \Omega)$; inferred probability that visitor n is in cognitive-style segment r ;
- $q_o(r_n)$ = prior cognitive-style segment probabilities;
- r_n = indexes cognitive-style segments, $r_n = 0$ to 15;
- $\vec{r}_n$ = vector notation for r_n as used in $\vec{\omega}_{r_n} = \Omega \vec{r}_n$; $\vec{r}_n$ is coded as four binary indicators plus a constant;
- $R(\alpha_{rmn}, \beta_{rmn}, a)$ = expected reward for acting optimally conditioned on $\alpha_{rmn}, \beta_{rmn}$, and a as used in the Bellman equation;
- s_{tn} = paired-comparison selection for the t th conjoint question for the n th priming visitor;
- t indexes the constant-sum questions; $t = 1$ to 8;

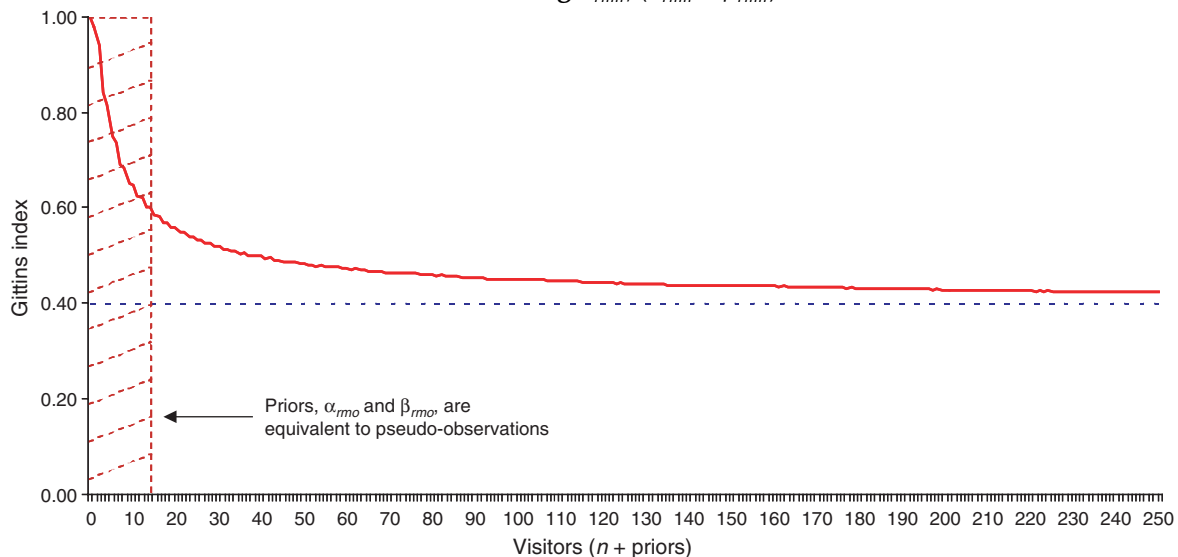
$\tilde{u}_{k|n}$ = visitor n 's utility for the j th click-alternative of the k th click; implies clickstream likelihood;
 $y_{k|n}$ = 1 if visitor n chooses the j th click-alternative on the k th click, 0 otherwise;
 $\tilde{y}_{kn}$ = binary vector for the k th decision point for the n th visitor;
 $\tilde{y}_n$ = clickstream matrix for the n th visitor;
 $\tilde{y}$ = set of $\tilde{y}_n$ s for all n , used only in summary notation;
 α_{rmn} = parameter of the naturally conjugate beta distribution used in the Gittins dynamic program (α_{rmo} is a prior value);
 β_{rmn} = parameter of the naturally conjugate beta distribution used in the Gittins dynamic program (β_{rmo} is a prior value);

δ_{mn} = indicator variable to indicate when the n th visitor purchases a BT broadband plan after seeing morph, m ; δ_{rmn} when r is known and we wish to make dependence on r explicit;
 $\tilde{\delta}$ = matrix of the δ_{mn} s, used in summary notation only;
 $\tilde{\epsilon}_{k|n}$ = extreme-value errors for choice among click-alternatives;
 γ = scaling parameter to allow scale differences in clickstream and paired-comparison data;
 $\tilde{\omega}_{r_n}$ = preference vector for the r_n th cognitive-style segment; used in $\tilde{u}_{k|n} = \tilde{c}'_{k|n} \tilde{\omega}_{r_n} + \tilde{\epsilon}_{k|n}$;
 Ω = matrix of the $\tilde{\omega}_{r_n}$; Ω is a 10×5 matrix;
 $\tilde{\xi}_{tn}$ = extreme-value measurement error used for paired-comparison conjoint questions.

Appendix 2. Quantiles of Posterior Distribution of Ω

Quantile	Mean effect		Leader versus follower		Analytic/verbal versus verbal/holistic		Impulsive versus deliberative		Reader versus listener	
	5%	95%	5%	95%	5%	95%	5%	95%	5%	95%
Graphical versus verbal	1.58	2.05	−0.13	0.16	−0.12	0.18	−0.15	0.14	−0.24	0.04
Small versus large load	−2.08	−1.63	−0.09	0.22	−0.28	0.06	−0.01	0.31	−0.16	0.14
Focused versus general	−0.28	0.12	−0.27	0.09	−1.03	−0.69	−0.21	0.12	0.11	0.44
General information	−0.11	0.26	−0.25	0.10	−0.54	−0.22	−0.24	0.09	−0.07	0.26
Analytic tool	0.94	1.19	−0.18	0.06	−0.11	0.13	−0.17	0.05	−0.14	0.09
Read a post	3.10	3.74	−0.42	0.10	−0.26	0.32	−0.15	0.32	−0.30	0.19
Post a comment	0.33	0.69	−0.20	0.16	−0.05	0.31	−0.22	0.15	−0.31	0.05
Compare plans	2.31	2.84	−0.41	0.12	0.43	0.90	−0.24	0.20	−0.40	0.09
Virtual advisor	1.34	1.90	−0.39	0.13	0.03	0.51	−0.36	0.11	−0.31	0.18
Community	—	—	—	—	—	—	—	—	—	—
Learning center	−0.19	0.47	−0.58	0.03	−0.24	0.40	−0.33	0.26	−0.19	0.40

Appendix 3. Gittins Index as Function of n Holding $\alpha_{rmn}/(\alpha_{rmn} + \beta_{rmn}) = 0.40$



References

- Allinson, C. W., J. Hayes. 1996. The cognitive style index: A measure of intuition-analysis for organizational research. *J. Management Stud.* **33**(1) 119–135.
- Ansari, A., C. F. Mela. 2003. E-customization. *J. Marketing Res.* **40**(2) 131–145.
- Bergkvist, L., J. R. Rossiter. 2007. The predictive validity of multiple-item versus single-item measures of the same constructs. *J. Marketing Res.* **64**(2) 175–184.
- Bertsimas, D., J. Niño-Mora. 1996. Conservation laws, extended polymatroids and multi-armed bandit problems. *Math. Oper. Res.* **21**(2) 257–306.
- Bidel, S. L. L., F. Piat, T. Artières, P. Gallinari. 2003. Statistical machine learning for tracking hypermedia user behavior. S. A. Macskassy, R. Wilkinson, A. Goker, M. Bauer, eds. *UM'03—9th Internat. Conf. User Modeling: Second Workshop on Machine Learn., Inform. Retrieval and User Model.*, Johnstown, PA, 48–65.
- Braun, M., G. L. Urban, C. Lee, J. R. Hauser. 2008. Matching website characteristics to cognitive styles. Working paper, MIT Sloan School of Management, Cambridge, MA.
- Chickering, D. M., T. Paek. 2007. Personalizing influence diagrams: Applying online learning strategies to dialogue management. *User Model. User-Adapted Interaction* **17**(1–2) 71–91.
- Conati, C., A. Gertner, K. VanLehn. 2002. Using Bayesian networks to manage uncertainty in student modeling. *User Model. User-Adapted Interaction* **12**(4) 371–417.
- Drolet, A. L., D. G. Morrison. 2001. Do we really need multiple-item measures in service research? *J. Service Res.* **3**(3) 196–204.
- Dusonchet, F., M.-O. Hongler. 2006. Priority index heuristic for multi-armed bandit problems with set-up costs and/or set-up time delays. *Internat. J. Comput. Integrated Manufacturing* **19**(3) 210–219.
- Frederick, S. 2005. Cognitive reflection and decision making. *J. Econom. Perspectives* **19**(4) 25–42.
- Frias-Martinez, E., S. Y. Chen, X. Liu. 2007. Automatic cognitive style identification of digital library users for personalization. *J. Amer. Soc. Inform. Sci. Tech.* **58**(2) 237–251.
- Gittins, J. C. 1979. Bandit processes and dynamic allocation indices. *J. Roy. Statist. Soc. Ser. B* **41**(2) 148–177, plus commentary.
- Gittins, J. C. 1989. *Allocation Indices for Multi-Armed Bandits*. John Wiley & Sons, London.
- Harvey, O. J., D. E. Hunt, H. M. Schroder. 1961. *Conceptual Systems and Personality Organization*. John Wiley & Sons, New York.
- Hauser, J. R. 1978. Testing the accuracy, usefulness and significance of probabilistic models: An information theoretic approach. *Oper. Res.* **26**(3) 406–421.
- Hauser, J. R., O. Toubia. 2005. The impact of utility balance and endogeneity in conjoint analysis. *Marketing Sci.* **24**(3) 498–507.
- Hayes, J., C. W. Allinson. 1998. Cognitive style and the theory and practice of individual and collective learning in organizations. *Human Relations* **31**(7) 847–871.
- Hofstede, G. 1983. National cultures in four dimensions. *Internat. Stud. Management Organ.* **13** 46–74.
- Hofstede, G. 1984. *Culture's Consequences: International Differences in Work-Related Values*. Sage, Beverly Hills, CA.
- Hofstede, G. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*, 2nd ed. Sage Publications, Thousand Oaks, CA.
- Hughes, M. A., D. E. Garrett. 1990. Intercoder reliability estimation approaches in marketing: A generalizability theory framework for quantitative data. *J. Marketing Res.* **27**(May) 185–195.
- Jamieson, L. F., F. M. Bass. 1989. Adjusting stated intention measures to predict trial purchase of new products: A comparison of models and methods. *J. Marketing Res.* **26**(August) 336–345.
- Jun, T. 2004. A survey on the bandit problem with switching costs. *de Economist* **152**(4) 513–541.
- Kalwani, M. U., A. J. Silk. 1982. On the reliability and predictive validity of purchase intention measures. *Marketing Sci.* **1**(3) 243–286.
- Kirton, M. J. 1987. *Adaption-Innovation Inventory (KAI) Manual*. Occupational Research Centre, Hatfield, UK.
- Kopfstein, D. 1973. Risk-taking behavior and cognitive style. *Child Development* **44**(1) 190–192.
- Krishnamurthy, V., J. Mickova. 1999. Finite dimensional algorithms for the hidden Markov model multi-armed bandit problem. *IEEE Internat. Conf. Acoustics, Speech, and Signal Processing*, Vol. 5. Institute of Electrical and Electronics Engineers, Washington, D.C., 2865–2868.
- Liechty, J., R. Peters, M. Wedel. 2003. Global and local covert visual attention: Evidence from a Bayesian hidden Markov model. *Psychometrika* **68**(4) 519–541.
- Magoulas, G. D., K. A. Papanikolaou, M. Grigoriadou. 2001. Neuro-fuzzy synergism for planning the content in a Web-based course. *Informatica* **25** 39–48.
- Mainemelis, C., R. Boyatzis, D. A. Kolb. 2002. Learning styles and adaptive flexibility: Testing experiential learning theory. *Management Learn.* **31**(1) 5–33.
- Montgomery, A. L., S. Li, K. Srinivasan, J. Liechty. 2004. Modeling online browsing and path analysis using clickstream data. *Marketing Sci.* **23**(4) 579–585.
- Montoya, R., O. Netzer, K. Jedidi. 2008. Dynamic marketing resource allocation for long-term profitability: A pharmaceutical application. Working paper, Columbia University, New York.
- Morrison, D. G. 1979. Purchase intentions and purchase behavior. *J. Marketing* **43**(Spring) 65–74.
- Morwitz, V. G., E. Johnson, D. Schmittlein. 1993. Does measuring intent change behavior? *J. Consumer Res.* **20**(June) 46–61.
- Netzer, O., J. M. Lattin, V. Srinivasan. 2008. A hidden Markov model of customer relationship dynamics. *Marketing Sci.* **27**(2) 185–204.
- Paivio, A. 1971. *Imagery and Verbal Processes*. Holt, Rinehart and Winston, New York.
- Perreault, W. D., Jr., L. E. Leigh. 1989. Reliability of nominal data based on qualitative judgments. *J. Marketing Res.* **26**(May) 135–148.
- Rhoads, B., G. L. Urban, F. Sultan. 2004. Building customer trust through adaptive site design. *MSI Conference, Yale University, New Haven, CT, December 11*; and abstracted in Urban, Glen 2005. *Don't Just Relate—Advocate*. Pearson Education, Upper Saddle River, NJ, 146–148.
- Riding, R. J., I. Calvey. 1981. The assessment of verbal-imagery learning styles and their effect on the recall of concrete and abstract prose passages by eleven year old children. *British J. Psych.* **72**(1) 59–64.
- Riding, R. J., I. Cheema. 1991. Cognitive styles: An overview and integration. *Educational Psych.* **11**(3&4) 193–215.
- Riding, R. J., S. Rayner. 1998. *Cognitive Styles and Learning Strategies: Understanding Style Differences in Learning and Behavior*. David Fulton Publishers, London.
- Riding, R. J., E. M. Taylor. 1976. Imagery performance and prose comprehension in seven-year-old children. *Educational Stud.* **2**(1) 21–27.
- Rogers, E. M. 1962. *The Diffusion of Innovations*. The Free Press, New York.
- Rogers, E. M., J. D. Stanfield. 1968. Adoption and diffusion of new products: Emerging generalizations and hypotheses. F. M. Bass, C. W. King, E. A. Pessimer, eds. *Applications of the Sciences in Marketing Management*. John Wiley & Sons, New York, 227–250.

- Rust, R. T., B. Cooil. 1994. Reliability measures for qualitative data: Theory and implications. *J. Marketing Res.* **31**(February) 1–14.
- Santally, M. I., S. Alain. 2006. Personalisation in Web-based learning environments. *J. Distance Ed. Tech.* **4**(4) 15–35.
- Siegelman, E. 1969. Reflective and impulsive observing behavior. *Child Development* **40**(4) 1213–1222.
- Steenkamp, J.-B. E. M., H. Baumgartner. 1998. Assessing measurement invariance in cross-national consumer research. *J. Consumer Res.* **25**(1) 78–90.
- Tarpin-Bernard, F., F. Habieb-Mammar. 2005. Modeling elementary cognitive abilities for adaptive hypermedia presentation. *User Modeling User-Adapted Interaction* **15**(5) 459–495.
- Train, K. 2003. *Discrete Choice Methods with Simulations*. University Press, Cambridge, UK.
- Trompenaars, F., C. Hampden-Turner. 1997. *Riding the Waves of Culture: Understanding Cultural Diversity in Business*. Nicholas Brealey, London.
- Tsitsiklis, J. N. 1994. A short proof of the Gittins index theorem. *Ann. Appl. Probab.* **4**(1) 194–199.
- Tybout, A. M., J. R. Hauser. 1981. A marketing audit using a conceptual model of consumer behavior: Application and evaluation. *J. Marketing* **45**(3) 82–101.
- Urban, G. L., J. R. Hauser. 2004. “Listening-in” to find and explore new combinations of customer needs. *J. Marketing* **68**(April) 72–87.
- Vermorel, J., M. Mohri. 2005. Multi-armed bandit algorithms and empirical evaluation. J. Gama, R. Camacho, P. Brazdil, A. Jorge, L. Torgo, eds. *Machine Learning: ECML 2005—Proc. Eur. Conf. Machine Learn. Porto, Portugal, October 3–7. Lecture Notes in Computer Science*, Vol. 3720. Springer, Berlin, 437–448.
- von Hippel, E. 1988. *Sources of Innovation*. Oxford University Press, New York.
- Weber, R. 1992. On the Gittins index for multi-armed bandits. *Ann. Appl. Probab.* **2**(4) 1024–1033.
- Wernerfelt, B. 1996. Efficient marketing communications: Helping the customer learn. *J. Marketing Res.* **23**(2) 239–243.
- Whittle, P. 1989. Foreword to *Multi-Armed Bandit Allocation Indices*, by J. C. Gittins. John Wiley & Sons, London.
- Witkin, H. A., C. Moore, D. Goodenough, P. Cox. 1977. Field-dependent and field-independent cognitive styles and their educational implications. *Rev. Educational Res.* **47**(1) 1–64.
- Wright, P. 1973. The cognitive processes mediating acceptance of advertising. *J. Marketing Res.* **10**(February) 53–62.
- Yudelson, M. V., O. P. Medvedeva, R. S. Crowley. 2008. A multifactor approach to student model evaluation. *User Model. User-Adapted Interaction* **18**(4) 349–382.