Website Morphing

by

John R. Hauser

Glen L. Urban

Guilherme Liberali

and

Michael Braun

August 2008

John R. Hauser is the Kirin Professor of Marketing, MIT Sloan School of Management, Massachusetts Institute of Technology, E40-179, 1 Amherst Street, Cambridge, MA 02142, (617) 253-2929, fax (617) 253-7597, jhauser@mit.edu.

Glen L. Urban is the David Austin Professor of Marketing, MIT Sloan School of Management, Massachusetts Institute of Technology, E40-159, 1 Amherst Street, Cambridge, MA 02142, (617) 253-6615, fax (617) 258-7597, glurban@mit.edu.

Guilherme (Gui) Liberali is a Visiting Scholar at MIT Sloan School of Management, and an Assistant Professor of Marketing, Universidade do Vale do Rio dos Sinos, Av.UNISINOS 950, Sao Leopoldo,RS 90450 – Brazil, (55.51)3590-8186, fax (55.51)3590-8447, liberali@unisinos.br.

Michael Braun is an Assistant Professor of Marketing, MIT Sloan School of Management, Massachusetts Institute of Technology, E40-169, 1 Amherst Street, Cambridge, MA 02142, (617) 253-3436, fax (617) 253-7597, braunm@mit.edu.

Website Morphing

Abstract

Virtual advisors often increase sales for those customers who find such on-line 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 click-stream 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.

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 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 dell.com), recommendations (as in amazon.com), factorial experiments (Google's website optimizer), or customized content (Ansari and Mela 2003, Montgomery, Li, Srinivasan, and Liechty 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) vs. deliberative (explores options in depth before making a decision),

---

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

visual (prefers images) vs. verbal (prefers text and numbers), or analytic (wants all details) vs. holistic (just the bottom line). (We provide greater detail and citations in a later section.) 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. We use 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, Moore, Goodenough and Cox 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 vs. holistic, and are usually measured either by question banks or cognitive tasks (Frias-Martinez, Chen and Liu 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 (Section 7, BT application).

Figure 1 illustrates two of the eight versions ("morphs") of broadband advisors from the BT application. Figure 1a uses an analytic virtual advisor (a technology magazine editor willing to provide data) who compares plans on ten characteristics (a large information load), presents a bar chart to compare prices (graphical), and displays a complete comparison across all plans (general content). In contrast, Figure 1b 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 (focused content).

Figure 1
Comparison of Two Morphs for a Website Advisor
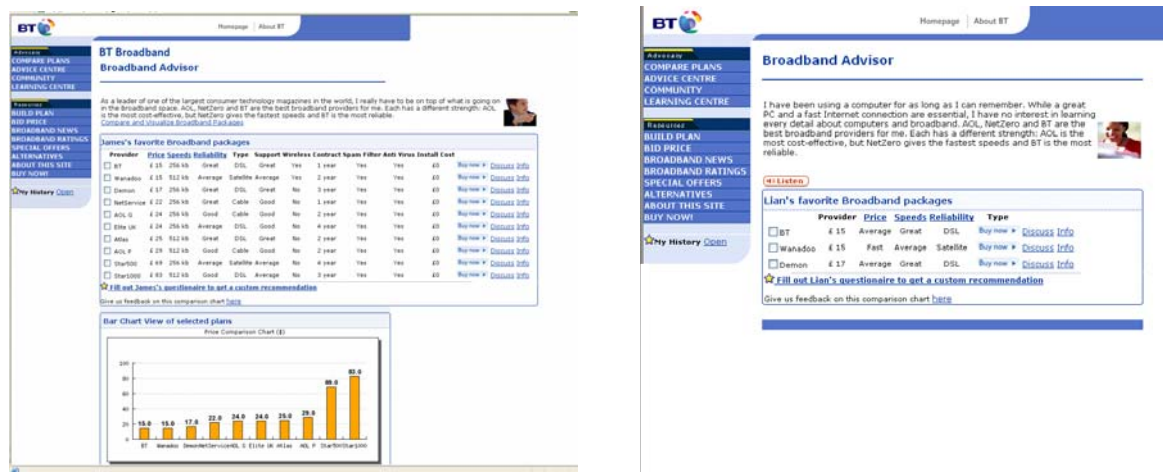
(a) General content, large-load, graphical morph   (b) Focused, small-load, verbal morph

We expect different morphs to appeal differentially depending upon visitors' cognitive styles. For example, impulsive visitors might prefer less-detailed information while 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 Section 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 nth 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. If we had perfect information on cognitive-style segments and perfect knowledge of segment x morph purchase probabilities, we could map an optimal morph to each cognitive-style segment. There are 16 x 8 = 128 such segment x 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 ten), 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 either 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 Section 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 nth visitor faces on the kth click. Let y_{kjn} be 1 if the nth visitor chooses the jth alternative on the kth click, and 0 otherwise. Let $\vec{y}_{kn}$ be the vector of the y_{kjn}'s and let $\vec{y}_n$ be the matrix of the $\vec{y}_{kn}$'s. Each click-alternative is described by a set of characteristics, $\vec{c}_{kjn}$. In our application, there are eleven characteristics: three macro characteristics (e.g., visual vs. 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 charac-

teristics to other combinations. Let $\vec{\omega}_{r_n}$ be a vector of preference weights that maps click-alternative characteristics, $\vec{c}_{kjn}$, to preference for each cognitive-style segment, r_n. Define Ω as the matrix of the $\vec{\omega}_r$'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/she makes a decision to click), and (3) the clicks that were made, we can infer the visitor's cognitive-style segment with Bayes' Theorem. Specifically, we update the posterior distribution, $f(r_n | \vec{y}_n, \Omega, \vec{c}_{kjn}'s)$, that the visitor is in the r_n^{th} segment based on the observed data.[2]

Figure 2
Website Morphing: Bayesian Inference Loop (grey dashed line) and Dynamic-Programming Inference/Optimization Loop (black dotted line)
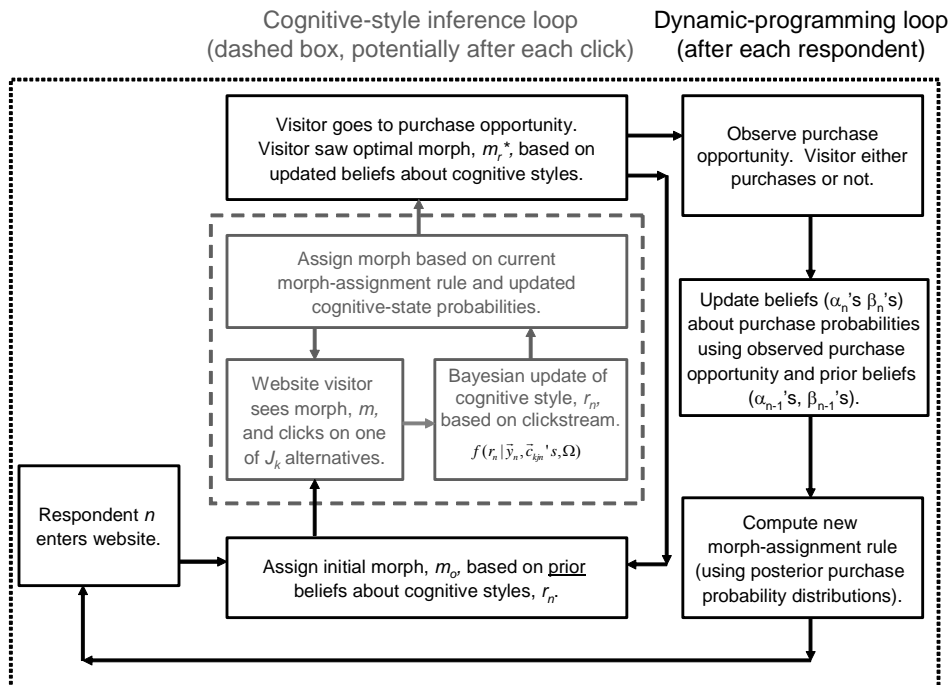
The second inference loop (outer loop denoted by a black dotted line in Figure 2) identifies the optimal morph conditioned on $f(r_n | \vec{y}_n, \Omega, \vec{c}_{kjn}'s)$. This inference loop must learn and optimize simultaneously. In theory, we might allow the website to morph many times for each visitor, 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

---

[2] This posterior distribution depends upon the morph, m_o, that the nth visitor has experienced. We have suppressed this subscript for ease of exposition. We explore estimation of Ω in Section 8.

characteristics, our initial application morphs only once per visitor. (We address multiple morphs later.) 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 $\vec{p}$ be the matrix of the p_{rm}'s. Clearly, if we knew r_n and the $\vec{p}$ perfectly, then we would assign the morph that maximizes p_{rm}. However, we do not know either r_n or $\vec{p}$ perfectly; we have only posterior probabilistic beliefs about r_n and $\vec{p}$. 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 ("multi-armed 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 review briefly 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 mate-

rial for distance learning, web-based learning, digital libraries, and hypermedia navigation. In most cases cognitive styles are measured with either an intensive inventory of psychometric scales or inferred from pre-defined tasks (Frias-Martinez, Chen and Liu 2007; Magoulas, Papanikoaou and Grigoriadou 2001; Mainemelis, Boyatzis and Kolb 2002; Santally and Alain 2006; Tarpin-Bernard and Habieb-Mamar 2005). Methods include direct classification, neuro-fuzzy logic, decision trees, multi-layer 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 a 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 multi-armed 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, Gertner and Vanlehn (2002) identify students' mastery of Newton's laws by pre-defining 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, Medvedeva and Crowley (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, Pieters and Wedel (2003) identify visual attention levels for advertising; and Netzer, Lattin and Srinivasan (2007) identify customer attitudes for alumni gift giving. Estimation methods include machine learning and hierarchical-Bayes Monte-Carlo-Markov-Chain methods. Montoya, Jedidi and Netzer (2007) estimate a 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 on-line 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 Section 7 and Supplemental Appendix on How to Build a Morph Taxonomy). 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 <u>temporarily</u> 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 nth 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, $f(\vec{p} \mid \delta_{mn}, parameters\ based\ on\ previous\ visitors)$.

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, $f(p_{rm} \mid \alpha_{rmo}, \beta_{rmo}) \sim$ $p_{rm}^{\alpha_{rmo}-1}(1 - p_{rm})^{\beta_{rmo}-1}$. With beta priors and Bernoulli observations, it is easy to show that the posterior is also a beta distribution with $\alpha_{rm,n+1} = \alpha_{rmn} + \delta_{mn}$ and $\beta_{rm,n+1} = \beta_{rmn} + (1 - \delta_{mn})$. 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 nth 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^*, that maximizes the sum of the expectation of the immediate reward and the discounted future reward.

In general, such "multi-arm 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 1979, p. ix). However, in a now-classic paper

Gittins (1979) proposed a simple and practical solution that decomposed the problem into indices. In 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 <u>known</u> 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 <u>strategy</u> 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, p. 8).

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: $\frac{G_{rmn}}{1-a}$. The reward for selecting an uncertain arm is more complicated to derive because each success or failure updates our beliefs about the probability of success.

Following standard dynamic programming notation we let R(α_{rmn}, β_{rmn}, a) be the value of acting optimally. To act optimally, we must choose one of two actions, either the known arm or the uncertain arm. When we select the uncertain arm we either get a success (with probability $\frac{\alpha_{rmn}}{\alpha_{rmn}+\beta_{rmn}}$) or a failure (with probability $\frac{\beta_{rmn}}{\alpha_{rmn}+\beta_{rmn}}$). 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, $\alpha_{rmn+1} = \alpha_{rmn} + 1$ and $\beta_{rmn+1} = \beta_{rmn}$. Thus, we expect a discounted reward of $1 + a$R($\alpha_{rmn}+1$, β_{rmn}) when we observe a success. By similar reasoning, we expect a discounted reward of aR(α_{rmn}, $\beta_{rmn}+1$) when we observe a failure. Putting

---

[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.

these rewards together we calculate the expected reward of uncertain arm as: $\frac{\alpha_{rmn}}{\alpha_{rmn}+\beta_{rmn}}[1+$

$a\text{R}(\alpha_{rmn}+1, \beta_{rmn}, a)] + \frac{\beta_{rmn}}{\alpha_{rmn}+\beta_{rmn}} a\text{R}(\alpha_{rmn}, \beta_{rmn}+1, a)$. Our strategy is to choose the arm with the

highest expected discounted profit, hence the Bellman equation becomes:

$$(1)\quad \text{R}(\alpha_{rmn}, \beta_{rmn}, a) = \max\left\{ \frac{G_{rmn}}{1-a}, \frac{\alpha_{rmn}}{\alpha_{rmn}+\beta_{rmn}}[1 + a\text{R}(\alpha_{rmn}+1, \beta_{rmn}, a)] + \frac{\beta_{rmn}}{\alpha_{rmn}+\beta_{rmn}} a\text{R}(\alpha_{rmn}, \beta_{rmn}+1, a)\right\}$$

Equation 1 has no analytic solution, but we can readily compute Gittins' indices with a simple iterative numeric algorithm.[4] We illustrate G_{rmn} 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_{rmn} exceeds $\frac{\alpha_{rmn}}{\alpha_{rmn}+\beta_{rmn}}$ substantially. As we observe more website visitors, G_{rmn} decreases as a function of n As $n \to \infty$ the expected rewards become known and G_{rmn} converges to $\frac{\alpha_{rmn}}{\alpha_{rmn}+\beta_{rmn}}$.   The discount rate, a, is constant for our application, but if a were to increase, we would value the future more and G_{rmn} would increase to make exploration more attractive.

Given a we pre-compute a table of indices for the values of α_{rmn} and β_{rmn} that we expect to observe in the BT application, using interpolation if necessary. The α-β table is made manageable by recognizing that G_{rmn} converges to $\frac{\alpha_{rmn}}{\alpha_{rmn}+\beta_{rmn}}$ as the number of visitors gets large.

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 1/100,000[th] 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 & 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 <u>if</u> we allow morphing. The question here is whether morphing

---

[4] We are indebted to Prof. John Gittins for sharing his code with us.

per se is reasonable. 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 8 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 x morph combination. (Details in Sections 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 an example cognitive-style segment. The first panel plots the evolution of the Gittins' indices; the second panel plots the morph chosen by the system. The Gittins' indices for each of the 8 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 1200th visitor. Around the 2500th visitor the system flirts with Morph 3 (cyan line), before settling down again on Morph 2. This blip around the 2500th visitor is due to 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.

---

[5] However, we would still have to be able to identify the no-morph strategy – itself a Gittins' problem.

Figure 3
Evolution of the Dynamic Programming Loop
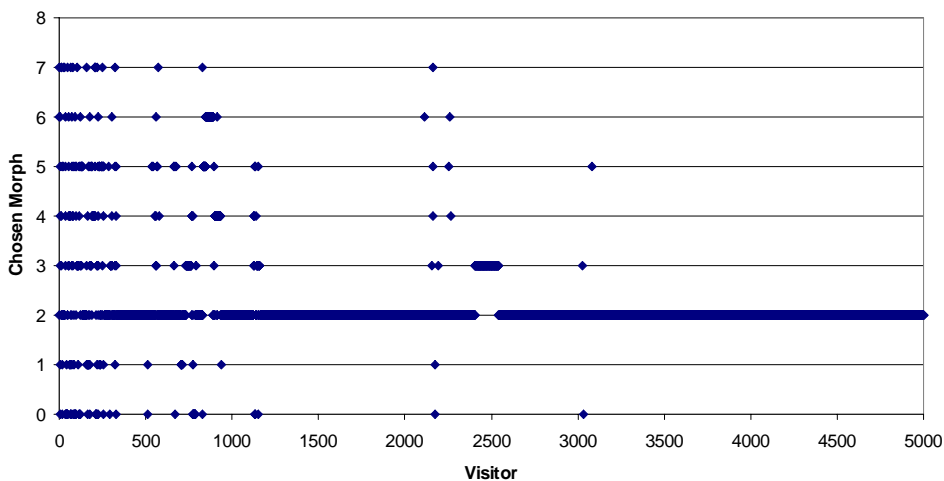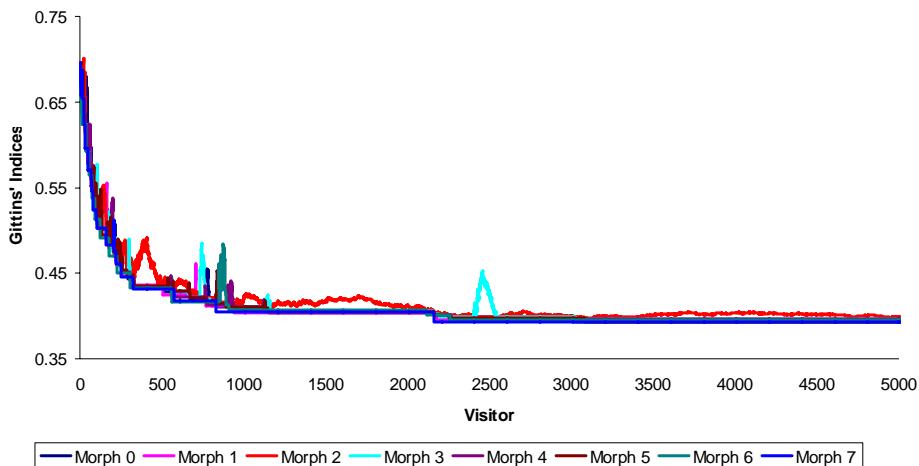
Because the Gittins' strategy is optimal in the presence of uncertainty, we can calculate the cost of uncertainty for this cognitive-style segment. The best morph for this segment is Morph 2 with an expected reward of 0.4023 times BT's profit per sale. If we had perfect information we would always choose 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 cogni-

12

tive styles in mind, is equivalent to one for which BT choses 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 (Section 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 Section 6 how the clickstream provides a posterior probability, $q_{rn} = f(r_n \mid \vec{y}_n, \Omega, \vec{c}_{kjn} \text{'} s)$, 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 partially-observable Markov decision process (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. While 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 rth 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:

$$(2) \qquad \alpha_{rmn} = \alpha_{rm,n-1} + \delta_{mn}\, q_{rn} \qquad\qquad \beta_{rmn} = \beta_{rm,n-1} + [1 - \delta_{mn}]\, q_{rn}$$

Second, following Krishnamurthy and Michova (KM, 1999) we compute an expected reward over the distribution of cognitive-style segments (the vector of probabilities q_{rn}) as well

---

[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 mth morph for the entire visit.

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, this simple heuristic solution, called an Expected Gittins' Index strategy [EGI], 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:

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

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 which would be obtained if BT were able to have perfect information on cognitive styles. The EGI solution does quite well (details in the next section).

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 get 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 $p_{r\{m1, m2, m3, ...\}}$ on a sequence of morphs, $\{m_1, m_2, m_3, ...\}$. 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

---

[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 endogenously using the EGI solution to the POMDP.

service. Thus, rather than asking website visitors to describe directly their cognitive styles, the Bayesian loop infers cognitive styles. Specifically, after observing the clickstream, $\vec{y}_n$, and the click-alternative characteristics, $\vec{c}_{kjn}$'s, we update the probabilities that the nth visitor belongs to each of the cognitive-style segments (q_{rn}'s). (Although the $\vec{c}_{kjn}$'s depend upon the initial morph, m_o, seen by the nth visitor, we continue to suppress the m_o subscript to keep the notation simple.)

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

$$(4) \qquad f(\vec{y}_{kn} \mid \vec{c}_{kjn}'s, r_n, \Omega) = \prod_{j=1}^{J_k} \left(\frac{\exp[\vec{c}'_{kjn}\vec{\omega}_{r_n}]}{\sum_{\ell=1}^{J_k} \exp[\vec{c}'_{k\ell n}\omega_{r_n}]} \right)^{y_{kjn}}$$

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

$$(5) \qquad q_{rn} = f(r_n \mid \vec{y}_n, \vec{c}_{kjn}'s, \Omega) = \frac{\prod_{k=1}^{K_n}\prod_{j=1}^{J_k} f(\vec{y}_{kn}\mid\vec{c}_{kjn}'s, r_n,\Omega)q_o(r_n)}{\prod_{k=1}^{K_n}\prod_{j=1}^{J_k}\sum_{r_o=0}^{15} f(\vec{y}_{kn}\mid\vec{c}_{kjn}'s, r_n,\Omega)q_o(r_n)}$$

where the $q_o(r_n)$ are the prior probabilities that the nth 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, 3GHz, 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, $\vec{c}_{kjn}$'s, are data. We obtain $q_o(r_n)$ and Ω from a priming study as described in Section 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 poste-

rior 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 due to 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.

The Effect of Imperfect Cognitive-Style Identification

      In Section 5 we found that the cost of uncertainty in segment x 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 morph x 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:

- 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.

---

[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 Section 4 are for a representative cognitive-style segment. The benchmarks cited here are based on the results of all sixteen cognitive-style segments.

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

- 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 webpages ($\vec{c}_{jkn}$'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_{rn} and choose an optimal morph based on the expected Gittins' indices (EGI). 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 EGI. (The EGI 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_{rn}'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 which 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 (Sections 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 Section 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 percent 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.

Table 1
Benchmark Comparison of the Inference and Optimization Systems

	Expected Reward	Improve- ment	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 either cognitive styles or purchase probabilities.

With 10 clicks the Bayesian loop can identify most cognitive states. The median posterior probability (q_{rn}) 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. With 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 Section 10.) A 20% increase in sales corresponds to approximately an $80M increase in revenue. The Gittins' loop projects a gain of approximately $52.3 M by finding the best morph even without customization. The 10-click Bayesian loop adds another $27.4 M by customizing the look and feel of the website based on posterior cognitive-style-segment probabilities. This is within

$2.6M of what could be obtained with 50 clicks. Perfect information on cognitive-style segments would add yet another $1.8 M bringing us to $84.1 M. 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 Sections 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_{rmo}'s , β_{rmo}'s, and $q_o(r_n)$'s and data with which to estimate preference weights (Ω) for website characteristics.

Priming Study – Questionnaires to Potential BT Website Visitors

Using a professional market research company (Applied Marketing Science, Inc.) and a respected British on-line 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 x 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 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 8 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.

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, a supplemental appendix, available from the authors (and the *Marketing Science* website), 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 vs. 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, Allison and Hayes, 1996, Kirton 1987, 1985 and Riding and Cheema 1991). Researchers in both psychology and marketing suggest that cognitive styles can be further differentiated as either impulsive or deliberative (Kopfstein 1973, and Siegelman 1969). With a slight rescaling three cognitive reflection scales developed by Frederick (2005) differentiate respondents on the impulsive vs. deliberative dimension.[13] Other scales measure visual vs. verbal styles, a key cognitive concept in psychology (Harvey et. al. 1961, Paivio 1971, Riding and Taylor, 1976 and Riding and Calvey 1981). This dimension is particularly relevant to website design where the tradeoff between pictures and text is an important design element. While leadership is not commonly a cognitive-style dimension

---

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

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 Stanfeld 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.

Figure 4
Example Measures of Cognitive Styles

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 re-duce the 13 scales (10 scales from Figure 4 plus the 3 impulsive-vs.-deliberate scales) to four cognitive dimensions. (See Braun, et. al. 2008 for greater detail on scale development and analy-sis.) For the BT data, impulsive vs. deliberative and leader vs. follower were measured with suf-ficient reliability (0.55 and 0.80, respectively), analytic vs. holistic and visual vs. verbal com-bined to a single construct (0.56 reliability). The analyses identified a fourth dimension: a single scale, reader vs. listener. We suspect that this reader vs. listener scale was driven by the nature of the broadband-service websites that often give respondents a choice of reading text/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 vs. follower
- analytic/visual vs. holistic/verbal
- impulsive vs. deliberative
- (active) reader vs. (passive) listener

Using median splits, we define $16 = 2$x2x2x2 cognitive-style segments.[14]

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 homepage of the experimental BT website gives the respondent a choice of four content areas. We expect visitors with different cognitive styles to vary on their desire to visit different content areas on their first click.

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:

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

---

[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.

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 extend 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.

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).

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 informational 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)

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

---

[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.

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

Figure 5
Broadband Advice Centre Home Page
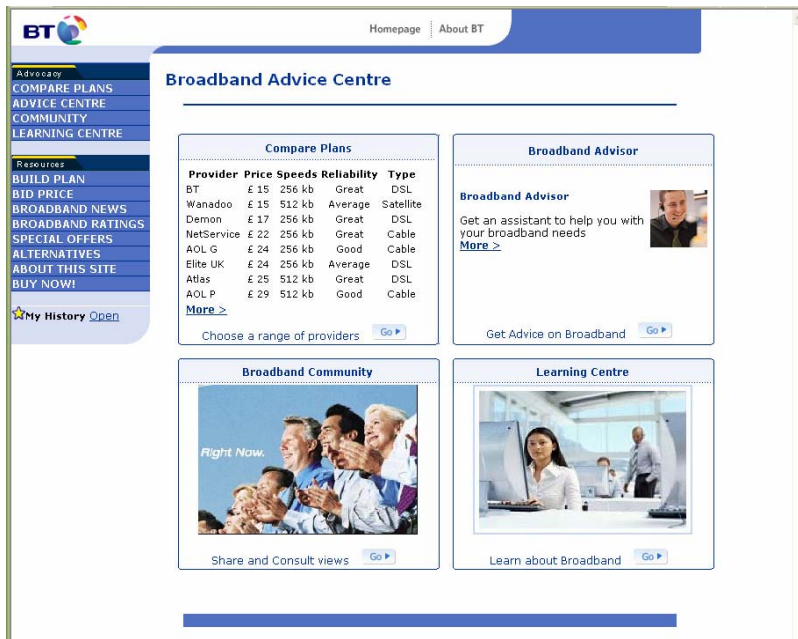
8. Estimation of Click-Alternative Preferences, Ω, From the Priming Data

The Bayesian inference loop uses visitors' clickstreams to compute posterior probabilities for cognitive-style segments r_n. The posterior probabilities (q_{rn}, Equation 5) require preference weights, Ω, for the click-alternative characteristics ($\vec{c}_{kjn}$'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 { $\vec{y}_n$'s, $\vec{c}_{kjn}$'s, r_n's }.

We have two sources of data within the priming study. First, we observe each respondent's clickstream. Second, we augment each respondent's clickstream 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.

Cognitive-style-segment Vector Notation

In Sections 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 5x1 binary vector, $\vec{r}_n$, where the first element is always equal to 1 and represents the characteristic-specific mean. Each subsequent element of $\vec{r}_n$ reflects a deviation from that mean based on a cognitive-style dimension of the segment. For example, a member of cognitive-style segment $r_n = 0 \Leftrightarrow \vec{r}_n'$ = {1,-1,-1, -1, -1} is a follower, holistic/verbal, deliberative, and a listener; $r_n = 15 \Leftrightarrow \vec{r}_n' =$ {1,1,1,1,1} is a leader, analytic/visual, impulsive, and a reader. With this notation, we write characteristic preferences compactly as $\vec{\omega}_{r_n} = \Omega \vec{r}_n$.

Clickstream likelihood

Using the vector notation combined with the notation of Sections 2-6, the clickstream likelihood (CSL) is based on Equation 5, except that Ω is unknown and the $\vec{r}_n$ are data. This likelihood assumes the unobserved errors are independent across clicks:

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

Paired-comparison likelihood

Each respondent is presented with eight pairs of website pages that vary on the three basic morph characteristics of graphic vs. verbal, focused vs. general, and small- vs. 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 for the tth 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

25

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:

$$(7) \qquad PCL = \prod_{n=1}^{835} \prod_{t=1}^{8} \left(\frac{\exp[\gamma\,\vec{d}_{tn1}'\Omega\vec{r}]}{\exp[\gamma\,\vec{d}_{tn1}'\Omega\vec{r}] + \exp[\gamma\,\vec{d}_{tn2}'\Omega\vec{r}]} \right)^{s_{tn}}$$

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]

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.

$$(8) \qquad f(\Omega, \gamma \mid \vec{c}_{kjm}, \vec{d}_{tn1}, \vec{d}_{tn2}, s_{tn}, \vec{y}_n, \vec{r}_n \; \forall\, k, j, t, n) \propto PCL * CSL * g(\Omega, \gamma)$$

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 are able to explain 60.3% of uncertainty in the clickstream choices ($U^2 = 0.603$, Hauser 1978).

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

---

[16] The standard deviations of the error terms, ε_{kjn} 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 (MLE). The Bayesian and MLE estimates were statistically equivalent.

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.

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

		Mean Effect	Leader vs. Follower	Analytic/visual vs. Holistic/verbal	Impulsive vs. Deliberative	Reader vs. listener
Basic Dimen-sions	Graphical vs. verbal	**1.82**	0.02	0.03	-0.01	-0.10
	Small vs. large load	**-1.85**	0.07	-0.11	**0.15**	-0.02
	Focused vs. general	-0.09	-0.09	**-0.86**	-0.04	**0.27**
Functional Character-istics	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

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. While 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.

---

[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 click-stream 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 hyper-distribution 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 x 16 segments x 10 clicks x 10 alternatives per click).

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 like 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 Section 4 suggest that we can achieve a slight improvement with stronger priors.

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.

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 x 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 $N_{rmo} \cong 12$.

Caveats and Practical Considerations

With sufficiently many website visitors from whom to observe actual purchase decisions, the $\bar{p}_{rmn}$ will converge to their true values and the priors will have negligible influence. None-

theless, 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 8 morphs, the average sample size is small for each segment x morph estimate of $\bar{p}_{rmo}$. To make more efficient use of the data and smooth these estimates over the r x 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 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 observe purchases.[20] Purchase intentions have the benefit of obtaining a more discriminating measure from each respondent than 0 vs. 1 purchase. However, purchase intentions are often subject to demand artifacts (e.g., Morwitz, Johnson and Schmittlein 1993). For example, for non-frequently-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 character-

---

[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.

istics implying an increase in revenue of $52.3 M (Table 1 and Section 6). Section 6 also suggests that a Gen-2 website (designed to distinguish among cognitive styles cleanly after ten clicks) could increase revenues an additional $27.4 M. 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.

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 vs. egalitarian, individual vs. collective, and emotional vs. neutral (Hofstede 1983, 1984; Trompenaars and Hampden-Turner 1997; and Steenkamp, Hofstede and Wedel 1999).

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 homepage 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 re-simulated 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 M) are much smaller than the potential gains with a Gen-2 website (total gains = $79.7M). Interestingly, even the Gen-1 website could get substantially more revenue if it had infinitely many visitors such that the system learned almost perfectly the segment x morph purchase probabilities (p_{rm}). Gen-1 ($n=\infty$) could achieve $75.7M 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, Li, Srinivasan, and Liechty 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 tradeoffs. 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. And, 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, p. 160-161) demonstrates that this normal-normal case is also solved with an index strategy and provides algorithms to the compute normal-normal indices. Vermorel and Mohri (2004) 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 <u>conditioned</u> 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., Dusonchet 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 were extended to other marketing-mix elements besides website design, we might consider latent states that evolved either randomly or based on marketing-mix elements. (See review in Section 3.) There are exciting opportunities to combine the advantages of HMMs or latent-class analysis with the exploration-exploitation tradeoffs made possible with expected Gittins' indices.

References

Allinson, Christopher W. and John Hayes (1996), "The Cognitive Style Index: A Measure of Intuition-Analysis for Organizational Research," *Journal of Management Studies*, 33, 1, (January), 119-135.

Ansari, Asim and Carl F. Mela (2003), "E-Customization," *Journal of Marketing Research*, 40, (May), 131-145.

Bergkvist, Lars and John R. Rossiter (2007), "The Predictive Validity of Multiple-Item Versus Single-Item Measures of the Same Constructs," *Journal of Marketing Research*, 64, 2, (May), 175-184.

Bertsimas, Dimitris and José Niňo-Mora (1996), "Conservation Laws, Extended Polymatroids And Multi-Armed Bandit Problems," *Mathematics of Operations Research*, 21, 2, (May), 257-306.

Bidel, S. L. Lemoine, F. Piat, T. Artières, and P. Gallinari (2003), "Statistical Machine Learning for Tracking Hypermedia User Behavior," *UM'03 - 9th International Conference on User Modeling: Workshop on Machine Learning, Information Retrieval and User Modeling.*

Biel, Alexander L. (1998), "Likeability: Why Advertising That Is Well Liked Sells Well," in *How Advertising Works: The Role of Research*, John Philip Jones, ed. (Thousand Oaks, CA: Sage Publications), 111-120.

Braun, Michael, Glen L. Urban, Clarence Lee, and John R. Hauser (2008), "Matching Website Characteristics to Cognitive Styles," Working Paper, MIT Sloan School of Management, Cambridge, MA 02142.

Chickering, David M. and Tim Paek (2007), "Personalizing Influence Diagrams: Applying Online Learning Strategies To Dialogue Management," *User Modeling and User-Adapted Interaction*, 17, 71-91.

Conati, Cristina and Abigail Gertner and Kurt Vanlehn (2002), "Using Bayesian Networks to Manage Uncertainty in Student Modeling," *User Modeling and User-Adapted Interaction*, 12, 371-417

Drolet, Aimee L. and Donald G. Morrison (2001), "Do we really need multiple-item measures in service research?" *Journal of Service Research*, 3, 3, (February), 196-204.

Dusonchet, Fabrice and Max-Olivier Hongler (2006), "Priority index heuristic for multi-armed bandit problems with set-up costs and/or set-up time delays," *International Journal of Computer Integrated Manufacturing*, 19, 3, (April-May), 210-219.

Frederick, Shane (2005), "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives*. 19(4). 25-42.

Frias-Martinez, Enrique, Sherry Y. Chen and Xiaohui Liu (2007), "Automatic Cognitive Style Identification of Digital Library Users for Personalization," *Journal of the American Society for Information Science and Technology,* 58, 2, 237-251.

Gittins, John C. (1979), "Bandit Processes and Dynamic Allocation Indices," *Journal of the Royal Statistical Society, Series B (Methodological)*, 41, 2, 148-177, plus commentary.

Gittins, John C. (1989), *Allocation Indices for Multi-Armed Bandits*, (London: Wiley).

Harvey, O.J., David E. Hunt, and Harold M. Schroder (1961), *Conceptual Systems and Personality Organization* (New York, NY: Wiley and Company).

Hayes, John and Christopher W. Allinson (1998), "Cognitive Style and the Theory and Practice of Individual and Collective Learning in Organizations," *Human Relations,* 31, 7, 847-71.

Hauser, John R. (1978), "Testing the Accuracy, Usefulness and Significance of Probabilistic Models: An Information Theoretic Approach," *Operations Research*, 26, 3, (May-June), 406-421.

----- and Olivier Toubia (2005), "The Impact of Utility Balance and Endogeneity in Conjoint Analysis," *Marketing Science*, 24, 3, (Summer), 498-507.

Hughes, Marie Adele and Dennis E. Garrett (1990), "Intercoder Reliability Estimation Approaches in Marketing: A Generalizability Theory Framework for Quantitative Data," *Journal of Marketing Research*, 27, (May), 185-195.

Hofstede, Geert (1983), "National Cultures in Four Dimensions," *International Studies of Management and Organization,*13, 46-74.

----- (1984), *Culture's consequences: International differences in Work Related Values,* (Beverly Hills, CA: Sage)

----- (2001), *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations, 2E* (Thousand Oaks, CA: Sage Publications, Inc.)

Jamieson, Linda F. and Frank M. Bass (1989), "Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods," *Journal of Marketing Research*, 26, (August), 336-345.

Jun, Tackseung (2004), "A Survey on the Bandit Problem with Switching Costs," *de Economist*, 152, 4, 513-541.

Kalwani, Manohar U. and Alvin J. Silk (1982), "On the Reliability and Predictive Validity of Purchase Intention Measures," *Marketing Science*, 1, 3, (Summer), 243-

Kirton, Michael J. (1987), *Adaption-Innovation Inventory (KAI) Manual* (Hatfield, UK: Occupational Research Centre).

Kopfstein Donald (1973), "Risk-Taking Behavior and Cognitive Style." *Child Development*, 44, 1, 190-192

Krishnamurthy, Vikram and Josipa Mickova (1999), "Finite Dimensional Algorithms For The Hidden Markov Model Multi-Armed Bandit Problem," *IEEE International Conference on Acoustics, Speech, and Signal Processing,* 5, 2865-2868.

Liechty, John, Rik Peters and Michel Wedel (2003), "Global and Local Covert Visual Attention: Evidence from a Bayesian Hidden Markov Model," *Psychometrika*, 68, 4, (December), 519-541.

Magoulas, G. D., K. A. Papanikolaou, and M. Grigoriadou (2001), "Neuro-fuzzy Synergism for Planning the Content in a Web-based Course," *Informatica*, 25, 39-48.

Mainemelis, Charalampos, Richard E. Boyatzis and David Kolb (2002), "Learning Styles and Adaptive Flexibility: Testing Experiential Learning Theory," *Management Learning,* 31, 1, (March), 5-33.

Montgomery, Alan L., Shibo Li, Kannan Srinivasan, and John Liechty (2004), "Modeling Online Browsing and Path Analysis Using Clickstream Data," *Marketing Science*, 23, 4, (Fall), 579-585.

Montoya, Ricardo, Kamel Jedidi and Oded Netzer (2007), "Dynamic Marketing Mix Allocation for Long-term Profitability," Working Paper, (New York, NY: Columbia University), December.

Morrison, Donald G. (1979), "Purchase Intentions and Purchase Behavior," *Journal of Marketing*, 43, (Spring), 65-74.

Morwitz, Vicki G., Eric Johnson, and David Schmittlein (1993), "Does Measuring Intent Change Behavior?," *Journal of Consumer Research*, 20, (June), 46-61.

Netzer, Oded, James Lattin, and V. Srinivasan (2007), "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science*, forthcoming.

Paivio, Allan (1971), *Imagery and Verbal Processes* (New York, NY: Holt, Rinehart & Winston, Inc.).

Perreault, William D., Jr. and Laurence E. Leigh (1989), "Reliability of Nominal Data Based on Qualitative Judgments," *Journal of Marketing Research*, 26, (May), 135-148.

Rhoads, Bryan, Glen L. Urban, and Fareena 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,* (Upper Saddle River, NJ: Pearson Education,) 146-48

Riding, Richard J. and 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 Journal of Psychology*, 72, 59-64

----- and Indra Cheema (1991), "Cognitive Style: An Overview and Integration", *Educational Psychology*, 11, 3&4, 193-215.

----- and Stephen Rayner (1998), *Cognitive Styles and Learning Strategies: Understanding Style Differences in Learning and Behavior*, (London, UK: David Fulton Publishers).

----- and E. M. Taylor (1976), "Imagery Performance And Prose Comprehension In 7 Year Old Children," *Educational Studies*, 2, 21-27

Rogers, Everett M. (1962), *The Diffusion of Innovations*, (New York, NY: The Free Press).

----- and J. David Stanfield, "Adoption and Diffusion of New Products: Emerging Generalizations and Hypotheses," A*pplications of the Sciences in Marketing Management*, Frank Bass, et. al., eds. (New York, NY: John Wiley & Sons, Inc.), 227-250.

Rust, Roland T. and Bruce Cooil (1994), "Reliability Measures for Qualitative Data: Theory and Implications," *Journal of Marketing Research*, 31, (February), 1-14.

Santally, Mohammad Issack and Seteni Alain (2006), "Personalisation in Web-Based Learning Environments," *Journal of Distance Education Technologies,* 4, 4, (October-December), 15-35.

Siegelman, Ellen (1969), "Reflective and Impulsive Observing Behavior" *Child Development*, 40, 4, 1213-222.

Steenkamp, Jan-Benedict E.M. and Hans Baumgartner (1998), "Assessing Measurement Invariance in Cross-National Consumer Research," *Journal of Consumer Research*, 25, 1, (June), 78-90.

Tarpin-Bernard,F and Habieb-Mammar, F. (2005). Modeling Elementary Cognitive Abilities for Adaptive Hypermedia Presentation, *User Modeling and User-Adapted Interaction*, 15, 459:495

Train, Kenneth (2003), "*Discrete Choice Methods with Simulations"*, (Cambridge, UK: University Press).

Tsitsiklis, John N. (1994), "A Short Proof of the Gittins Index Theorem," *Annals of Applied Probability*, 4, 1, (February), 194-199.

Trompenaars, Fons and Charles Hampden-Turner (1994), *Riding the Wales of Culture*: *Understanding Cultural Diversity in Business*, (London, UK: Nicholas Brealey).

Tybout, Alice M. and John R. Hauser (1981), "A Marketing Audit Using a Conceptual Model of Consumer Behavior: Application and Evaluation," *Journal of Marketing*, 45, 3, (Summer), 81-101.

Urban, Glen L. and John R. Hauser (2004), "'Listening-In' to Find and Explore New Combinations of Customer Needs," *Journal of Marketing*, 68, (April), 72-87.

Vermorel, Joannes and Mehryar Mohri (2005), "Multi-armed Bandit Algorithms and Empirical Evaluation," *Proceedings of the European Conference of Machine Learning,* Porto, Portugal (October 3-7). Springer Lecture Notes in Computer Science, vol. 3720, 437-448.

von Hippel, Eric (1988), *Sources of Innovation* (New York, NY: Oxford University Press).

Weber, Richard (1992), "On The Gittins Index For Multi-Armed Bandits," *Annals of Applied Probability*, 2, 4, (November), 1024–1033

Wernerfelt, Birger (1996), "Efficient Marketing Communications: Helping the Customer Learn," *Journal of Marketing Research*, 23, 239-243.

Whittle, Peter (1989), "Forward to Multi-armed Bandit Allocation Indices," by John Gittins, (London: Wiley).

Witkin, H. A., Moore, C., Goodenough, D., and Cox, P. (1977), "Field-dependent and Field-independent Cognitive Styles and their Educational Implications," *Review of Educational Research*, 47, 1-64.

Wright, Peter (1973), "The Cognitive Processes Mediating Acceptance of Advertising," *Journal of Marketing Research*, 10, (February), 53-62.

Yudelson, Michael, Olga Medvedeva and Rebecca Crowley (2008), "A Multifactor Approach to Student Model Evaluation," *User Modeling and User-Adapted Interaction*, Available Online as an Accepted Article in Advance.

Appendix 1. Notation (Optional, EIC to decide)

a = amount by which future visitors are valued, the discount rate in the dynamic program

$\vec{c}_{kjn}$ = characteristics of the jth click-alternative of the kth click decision by visitor n.

$\vec{d}_{tn1}$, $\vec{d}_{tn2}$ = first three elements of $\vec{c}_{kjn}$; notation used for paired-comparison selections

EG_{mn} = expected Gittins' index for m^{th} morph for visitor n

$f(\bullet)$ = probability density function, usually the posterior distribution

$g(\bullet)$ = probability density function, usually a prior

G_{rmn} = Gittins' index for r^{th} cognitive-style segment and m^{th} morph for visitor n

j indexes click-alternatives

J_{kn} = number of click-alternatives at the kth 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 m^{th} morph and are in the rth cognitive-style segment

o indices prior values, e.g., for α_{rmo}, β_{rmo}, p_{rmo}, N_{rmo}, m_o

p_{rm} = probability that visitor n in 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 nth visitor. $\bar{p}_{rmo}$ is mean of the prior for p_{rmo}

$\vec{p}$ = matrix of the p_{rm}'s

q_{rn} = $f(r_n | \vec{y}_n, \vec{c}_{kjn}'s, \Omega)$. 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.

$R(\alpha_{rmn}, \beta_{rmn}, a)$ = expected reward for acting optimally conditioned on α_{rmn}, β_{rmn}, and a as used in the Bellman equation.

s_{tn} = paired-comparison selection for the tth conjoint question for the nth priming visitor.

t indexes the constant-sum questions. $t = 1$ to 8.

$\tilde{u}_{kjn}$ = visitor n's utility for the jth click-alternative of the kth click; implies clickstream likelihood

y_{kjn} = 1 if visitor n chooses the jth click alternative on the kth click, 0 otherwise

$\vec{y}_{kn}$ = binary vector for the kth decision point for the nth visitor

$\vec{y}_n$ = clickstream matrix for the nth visitor

$\vec{y}$ = set of $\vec{y}_n$'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 nth 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.

$\vec{\delta}$ = matrix of the δ_{mn}'s, used in summary notation only

$\widetilde{\varepsilon}_{kjn}$ = extreme-value errors for choice among click-alternatives

γ = scaling parameter to allow scale differences in clickstream and paired-comparison data

$\vec{\omega}_{r_n}$ = preference vector for the r_n^{th} cognitive-style segment, used in $\widetilde{u}_{kjn} = \vec{c}'_{kjn}\vec{\omega}_{r_n} + \widetilde{\varepsilon}_{kjn}$.

Ω = matrix of the $\vec{\omega}_{r_n}$. Ω is a 10x5 matrix.

$\widetilde{\xi}_{tn}$ = extreme-value measurement error used for paired-comparison conjoint questions

Appendix 2. Quantiles of Posterior Distribution of Ω

	Mean Effect		Leader vs. Follower		Analytic/verbal vs. Verbal/holistic		Impulsive vs. Deliberative		Reader vs. listener	
Quantile	5%	95%	5%	95%	5%	95%	5%	95%	5%	95%
Graphical vs. verbal	1.58	2.05	-0.13	0.16	-0.12	0.18	-0.15	0.14	-0.24	0.04
Small vs. large load	-2.08	-1.63	-0.09	0.22	-0.28	0.06	-0.01	0.31	-0.16	0.14
Focused vs. 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 $\alpha_{rmn}/(\alpha_{rmn} + \beta_{rmn}) = 0.40$

priors, α_{rmo} and β_{rmo}, are equivalent to pseudo-observations

Gittins' Index

Visitors (n + priors)