MIT Open Access Articles

Morphing Banner Advertising

Citation	Urban, Glen L., Guilherme (Gui) Liberali, Erin MacDonald, Robert Bordley, and John R. Hauser. "Morphing Banner Advertising." Marketing Science 33, no. 1 (January 2014): 27–46.
As Published	http://dx.doi.org/10.1287/mksc.2013.0803
Publisher	Institute for Operations Research and the Management Sciences (INFORMS)
Version	Author's final manuscript
Citable link	http://hdl.handle.net/1721.1/90836
Terms of Use	Creative Commons Attribution-Noncommercial-Share Alike
Detailed Terms	http://creativecommons.org/licenses/by-nc-sa/4.0/

Morphing Banner Advertising

Abstract

Researchers and practitioners devote substantial effort to targeting banner advertisements to consumers, but focus less effort on how to communicate with consumers once targeted. Morphing enables a website to learn, automatically and near optimally, which banner advertisements to serve to consumers in order to maximize click-through rates, brand consideration, and purchase likelihood. Banners are matched to consumers based on posterior probabilities of latent segment membership, which are identified from consumers' clickstreams.

This paper describes the first large-sample random-assignment field test of banner morphing—over 100,000 consumers viewing over 450,000 banners on CNET.com. On relevant webpages, CNET's click-through rates almost doubled relative to control banners. We supplement the CNET field test with an experiment on an automotive information-and-recommendation website. The automotive experiment replaces automated learning with a longitudinal design that implements morph-to-segment matching. Banners matched to cognitive styles, as well as the stage of the consumer's buying process and body-type preference, significantly increase click-through rates, brand consideration, and purchase likelihood relative to a control. The CNET field test and automotive experiment demonstrate that matching banners to cognitive-style segments is feasible and provides significant benefits above and beyond traditional targeting. Improved banner effectiveness has strategic implications for allocations of budgets among media.

1. Introduction

This paper describes the first random-assignment field test of morphing with a sample size sufficient to observe steady-state behavior (116,168 unique CNET consumers receiving 451,524 banner advertisements). A banner advertisement morphs when it changes dynamically to match latent cognitive-style segments which, in turn, are inferred from consumers' clickstream choices. Examples of cognitive-style segments are impulsive-analytic, impulsive-holistic, deliberative-analytic, and deliberative-holistic. The website automatically determines the best "morph" by solving a dynamic program that balances exploration of morph-to-segment effectiveness with the exploitation of current knowledge about morph-to-segment effectiveness. Banner morphing modifies methods used in website morphing (Hauser, Urban, Liberali, and Braun 2009), which changes the look and feel of a website based on inferred cognitive styles. (For brevity we use HULB as a shortcut citation to the 2009 website-morphing paper.) Morphing adds behavioral-science-based dynamic changes which complement common banner-selection methods such as context matching and targeting.

HULB projected a 21% improvement in sales for BT Group's broadband-sales website, but the projections were based on simulated consumers whose behavior was estimated from data obtained *in vitro*. The BT Group did not allocate resources necessary to obtain sufficient sample for an *in vivo* field-test.[1] (By *in vivo* we refer to actual websites visited by real consumers for information search or purchasing. By *in vitro* we refer to laboratory-based websites that simulate actual websites and which are visited by a randomly recruited panel of consumers. *In vitro* experiments attempt to mimic *in vivo* field experiments, but never do so perfectly.)

Online morphing is designed for high-traffic websites with tens of thousands of visitors.

---

[1] Hauser, Urban, and Liberali (2012) report a field implementation of website-morphing with a small sample. Their results are suggestive but not significant. The morphing algorithm did not reach steady-state on their sample.

Simulations in HULB (Figure 3, p. 209) suggest that 10,000-20,000 consumers are necessary to realize substantial gains from website morphing. Banner morphing is likely to require higher sample sizes than website morphing because successful banner outcomes (click-throughs) occur relatively less often than successful website-morphing outcomes (sales of broadband services). Our field test (§4.9) has sufficient sample to observe a significant 83-97% lift in click-through rates between test and control cells above and beyond context matching.

Although click-through rates are a common industry metric, we also sought to test whether banner morphing increases brand consideration and purchase likelihood. Because brand-consideration and purchase-likelihood measures are intrusive, such metrics are difficult to obtain *in vivo*. We therefore supplement the large-sample field test with a smaller-sample random-assignment experiment on an *in vitro* automotive information-and-review website. We avoid the need for extremely large samples with three longitudinal surveys that act as surrogates for the HULB dynamic program. The first two surveys measure advertising preference, cognitive styles, and the stage of the consumer's buying process. The third survey, separated from the pre-measures by four and one-half (4½) weeks, exposes consumers to banner advertising while they search for information on cars and trucks. In the test group, consumers see banners that are matched to their cognitive style and buying stage. Banners are not matched in the control group. The sample (588 consumers) is sufficient because (1) we substitute direct measurement for Bayesian inference of segment membership and (2) we substitute measurement-based morph as-signment for the HULB dynamic program. The *in vitro* experiment suggests that matching banners to segments improves brand consideration and purchase likelihood relative to the control.

2. Banner Advertising – Current Practice

In the last ten years online advertising revenue has tripled. Banner advertisements, paid

advertisements placed on websites, account for 24% of online advertising revenue—about $6.2 billion in 2010. Banner advertisement placements cost roughly $10 per thousand impressions. Click-through rates are low and falling from 0.005 click-throughs per impression in 2001 to 0.001 in 2010 (Dahlen 2001; PricewaterhouseCoopers 2011). Website managers and marketing managers are highly interested in methods that improve banner effectiveness.

Current theory and practice attempt to increase click-through rates with a variety of methods. For example, Sundar and Kalyanaraman (2004) use laboratory methods to examine the effect of the speed and order of animation. Gatarski (2002) uses a genetic algorithm on a training sample to search 40 binary features of banners. He achieves a 66% lift above a 1% click-through rate based on sixteen "generations" seeing approximately 200,000 impressions.

Iyer, Soberman, and Villas-Boas (2005) and Kenny and Marshall (2000) suggest that click-through rates should improve when banners appear on webpages deemed to be relevant to consumers. Early attempts matched textual context. For example, Joshi, Bagherjeiran, and Ratnaparkhi (2011) cite an example where "divorce" in a banner is matched to "divorce" on the webpage. But context matters—it is not effective to place a banner for a divorce lawyer on a gossip account of a celebrity's divorce. Instead, Joshi, et al. achieved a 3.3% lift by matching a banner's textual context to a combination of webpage content and user characteristics. In a related application to Yahoo!'s news articles rather than banners, Chu, et al. (2009, p. 1103) use context-matching methods to increase click-through rates significantly (3.2% lift based on "several million page views"). Context matching is quite common. For example, General Motors pays Kelly Blue Book to show a banner advertisement for the Chevrolet Sonic when a consumer clicks on the compact-car category.

Relevance can also be inferred from past behavior: "[b]ehavioral targeting leverages his-

torical user behavior to select the most relevant ads to display." (Chen, Pavlov and Canny 2009, p. 209). Chen, et al. use cookie-based observation of 150,000 prior banners, webpages, and queries to identify the consumers who are most likely to respond to banners. They report expected lifts of approximately 16-26% based on in-sample analyses.

Laboratory experiments manipulate consumers' goals (surfing the web vs. seeking information) to demonstrate that banner characteristics, such as size and animation, are more or less effective depending upon consumers' goals (Li and Bukovac 1999; Stanaland and Tan 2010). This web-based research is related to classic advertising research that suggests advertising quality and endorser expertise (likability) are more or less effective depending upon relevance (involvement) for consumers (e.g., Chaiken 1980; Petty, Cacioppo and Schumann 1983).

Morphing differs from prior research in many ways. First, banners are matched to consumers based on cognitive styles rather than context relevance or past behavior. Second, latent cognitive-style segments are inferred automatically from the clickstream rather than manipulated in the laboratory. Third, morphing learns (near) optimally about morph-to-segment matches *in vivo* as consumers visit websites of their own accord. Thus, morphing is a complement rather than a substitute for existing methods such as context matching. If successful, morphing should provide incremental lift beyond context matching.

3. Brief Review of Banner Morphing

The basic strategy of morphing is to identify a consumer's segment from the consumer's clickstream and show that consumer the banner that is most effective for the consumer's segment. Because the clickstream data cannot completely eliminate uncertainty about the consumer's segment, we treat these segments as latent—we estimate probabilities of segment membership from the clickstream. In addition, there is uncertainty about which banner is most effective

for each latent segment. Using latent-segment probabilities and observations of outcomes, such as click throughs, the morphing algorithm learns automatically and near optimally which morph to give to each consumer. Morphing relies on fairly complex Bayesian updating and dynamic programming optimization. Before we provide those details, we begin with the conceptual description in Figure 1.

In Figure 1 we label the latent segments as Segment 1 through 4. Typically, the segments represent different cognitive styles, but segments can also be defined by other characteristics such as the stage of the consumer's buying process. A design team uses artistic skills, intuition, and past experience to design a variety of alternative websites (HULB) or alternative banners (this paper). We call these banners (or websites) "morphs." In Figure 1 we label the morphs as Morph 1 through Morph 4. Designers try to give the system a head start by designing morphs they believe match segments, but, *in vivo*, the best matches are identified automatically and optimally by the morphing algorithm.

[Insert Figure 1 about here.]

If the segments could be measured directly, rather than identified latently, the morphing optimization would be "indexable." Indexability implies we can solve the optimal allocation of morphs to segments by computing an index for each morph x segment combination. The index is called a Gittins' index. The Gittins' indices evolve based on observed consumers' behavior. The optimal policy for the nth consumer would be to assign the morph with the largest index for the consumer's segment. For example, if the upper-left bar chart represents the Gittins' indices computed after 100 consumers, and if segments were known, the algorithm would assign Morph 3 to Segment 4 because Morph 3 has the largest Gittins' index for Segment 4 (largest of the dark bars). Similarly, it would assign Morph 1 to Segment 2.

But segment membership cannot be observed directly. Instead the HULB algorithm uses a pre-calibrated Bayesian model to infer the probabilities that the consumer belongs to each latent segment. The probabilities are inferred from the clickstream on the website, possibly including multiple visits. Illustrative probabilities are shown by the bar chart in the middle of Figure 1. We use these segment-membership probabilities and the Gittins' indices to compute Expected Gittins' Indices (bar chart in the upper right of Figure 1). There is now one Expected Gittins' Index per morph. Based on research by Krishnamurthy and Mickova (1999), the (near) optimal policy for latent segments is to assign the morph with the highest Expected Gittins' Index. The bar chart in the upper-right corner tells us to assign Morph 3 to the 101st consumer. Because a sample size of 100 consumers is small, the system is still learning morph-to-segment assignments and, hence, the bars are more or less of equal height. If the 101st consumer had made different clicks on the website, the segment probabilities would have been different and, perhaps, the morph assignment would have been different.

As more consumers visit the website, we observe more outcomes—sales for website morphing or click-throughs for banner morphing. Using the observed outcomes the algorithm refines the morph x segment indices (details below). The middle-left and lower-left bar charts reflect refinements based on information up to and including the 20,000th and 80,000th consumer, respectively. As the indices become more refined, the morph assignments improve. (In Figure 1's illustrative example, the Expected Gittins' Index assigns Morph 3 after 100 consumers, changes to Morph 2 after 20,000, and discriminates even better after 80,000 consumers.)

State-of-the art morphing imposes limitations. First, because many observations are needed for each index to converge, the morphing algorithm is limited to a moderate number of morphs and segments. (HULB used $8 \, x \, 16 = 128$ Gittins' indices rather than the 16 indices in

Figure 1.) Second, although designers might create morphs using underlying characteristics and morphing may define segments based on underlying cognitive dimensions, the dynamic program does not exploit factorial representations. Schwartz (2012) and Scott (2010) propose an improvement to handle such factorial representations to identify the best banners for the non-morphing case, but their method has not been extended to morphing.

We now formalize the morphing algorithm. Our description is brief, but we provide full notation and equations in Appendix 1. Readers wishing to implement morphing will find sufficient detail in the cited references. Our code is available upon request.

3.1. Assigning Consumers to Latent Segments based on Clickstream Data

Figure 2 summarizes the two phases of morphing. We call the first phase a calibration study. The *in vitro* calibration study measures cognitive styles directly using established scales. Such measurement is intrusive and would not be feasible *in vivo*. Respondents for the calibration study are drawn from the target population and compensated to complete the calibration tasks. Using the questions designed to identify segment membership, we assign calibration-study consumers to segments. For example, HULB asked 835 broadband consumers to complete a survey in which the consumers answered 13 agree-vs.-disagree questions such as "I prefer to read text rather than listen to a lecture." HULB factor analyzed answers to the questions to identify four bipolar cognitive-style dimensions. They used median splits on the dimensions to identify sixteen (2 x 2 x 2 x 2 = 16) segments.

[Insert Figure 2 about here.]

Calibration study respondents explore an *in vitro* website as they would *in vivo*. We observe their chosen clickstream. We record each respondent's clickstream as well as the characteristics of all possible click choices (links) on the website. An example "click characteristic" is

7

whether the click promises to lead to pictures or text. Other click characteristics are dummy variables for areas of webpage (such as a comparison tool), expectations (the click is expected to lead to an overall recommendation), or other descriptions. These calibration data are used to estimate a logit model that maps click characteristics to the chosen clicks (see Appendix 1, Equation A1). The parameters of the logit model are conditioned on consumers' segments. The calibration study also provides the (unconditioned) percent of consumers in each segment—data that form prior beliefs for *in vivo* Bayesian calculations.

During day-to-day operation of the *in vivo* website we do not observe consumers' segments; instead, we observe consumers' clickstreams. The calibrated model and observed click characteristics give likelihoods for the observed clickstream conditioned upon a consumer belonging to each of the (now latent) segments. Using Bayes Theorem (and prior beliefs) we compute the probabilities that a consumer with the observed clickstream belongs to each segment (as shown in the middle of Figure 1.) See Appendix 1, Equation A1. In notation, let n index consumers, r index segments, and t index clicks. Let $\vec{c}_{nt}$ be consumer n's clickstream up to the tth click. The outcomes of the Bayesian calculations are the probabilities, $\Pr(r_n = r | \vec{c}_{nt})$, that consumer n belongs to segment r conditioned on the consumer's clickstream.

In HULB the first 10 clicks on the *in vivo* website were used to identify the consumer's segment and select the best morph. We adopt the same strategy of morphing after a fixed and pre-determined number of clicks. We label the fixed number of clicks with t_o. Hauser, Urban and Liberali (2012) propose a more complex algorithm to determine the optimal time to morph, but their algorithm was not available for our experiments. Thus, our experiments are conservative because morphing would likely do even better with improved algorithm.

3.2. Automatically Learning the Best Banner for Each Consumer

For ease of exposition, temporarily assume we can observe directly the consumer's latent segment. Let m index morphs and let p_{rm} be the probability of a good outcome (a sale or a click-through) given that a consumer in segment r experienced morph m for all clicks after the first t_o clicks. One suboptimal method to estimate p_{rm} would be to observe outcomes after assigning morphs randomly to a large number, N_{large}, of consumers. This policy, similar to that used by Google's web optimizer and many behavioral-targeting and context-matching algorithms, is sub-optimal during the calibration period because N_{large} consumers experience morphs that may not lead to the best outcomes.[2] To get a feel for N_{large}, assume eight morphs and four segments as in the CNET experiment, assume a typical click-through rate of 2/10$^{\text{ths}}$ of 1 percent, and calculate the sample size necessary to distinguish 2/10$^{\text{ths}}$ of 1 percent from a null hypothesis of 1/10$^{\text{th}}$ of 1 percent. We would need to assign suboptimal banners to approximately 128,000 consumers to obtain even a 0.05 level of significance (exact binomial calculations for each morph x segment). Morphing identifies optimal assignments with far fewer suboptimal banners.

Optimal assignment for directly observed segments is a classic problem in dynamic programming. The dynamic program balances the opportunity loss incurred while exploring new morph-to-segment assignments with the knowledge gained about the optimal policy. The updated knowledge is gained by observing outcomes (sales or click-throughs) and is summarized by posterior estimates of the p_{rm}'s. (See Appendix 1, Equation A2.) Improved posterior estimates enable us to assign morphs more effectively to future consumers.

For known segments, the optimal solution to the dynamic program has a simple form: we compute an index for each r, m-combination. The index, called the Gittins' index, G_{rmn}, is the

---

[2] Google is now implementing Gittins' experimentation but has not yet implemented morphing (private communication and http://support.google.com/analytics/bin/answer.py?hl=en&answer=2677320).

solution to a simpler dynamic program that depends only on assignments and outcomes for those consumers who experienced that r, m-combination (see Appendix 1, Equation A3). For the nth consumer, the optimal policy assigns the morph which has the largest index for the consumer's segment (Gittins (1979). The indices evolve with n.

Because we do not observe the consumer's segment directly, we must estimate the probabilities that the consumer belongs to each latent segment. Thus, *in vivo,* the problem becomes a partially observable Markov decision process (usually abbreviated POMDP). Krishnamurthy and Mickova (1999) establish that the POMDP is indexable and that an intuitive policy is near optimal. Their policy assigns the morph with the largest Expected Gittins' index. The Expected Gittins' Index is defined by $EG_{mn} = \sum_r \Pr(r_n = r \mid \vec{c}_{nt}) G_{rmn}$. We still update the p_{rm}'s and the G_{rmn}'s, but we now do so using the $\Pr(r_n = r | \vec{c}_{nt})$'s. The key differences between the Expected Gittins' Index policy and the naïve calibration-sample policy (N_{large}) is that the Expected Gittins' Index policy (1) learns while minimizing opportunity loss, (2) continues to learn as n gets large, and (3) can adapt when p_{rm} changes due to unobserved shocks such as changes in tastes, new product introductions, or competitive actions. Recalibration is automatic and optimal.

4. CNET Field Experiment

4.1. Smart Phone Banners on CNET.com

CNET.com is a high-volume website that provides news and reviews for high-tech products such as smart phones, computers, televisions, and digital cameras. It has 8 million visitors per day and has a total market valuation of $1.8 billion (Barr 2008). Banner advertising plays a major role in CNET's business model. Context-matched banners demand premium prices. For example, a computer manufacturer might purchase banner impressions on web pages that provide laptop reviews. Non-matched banners are priced lower. Morphing provides a means for

CNET to improve upon context-matching and, hence, provide higher value to its customers. CNET accepted our proposal to compare the performance of morphing versus a control on their website and to explore interactions with context matching.

The banners advertised AT&T smart phones. Consumers visiting CNET.com were assigned randomly to test and control cells. In each experimental cell some banners were context-matched and some were not (as occurred naturally on CNET). To assure sufficient sample for the morphing algorithm to be effective, we assigned 70% of the consumers to the test cell. CNET's agency developed a pool of eight AT&T banner advertisements about HTC refurbished smart phones. Five of the banners were square banners that could appear anywhere on the website and three of the banners were wide rectangular banners that appear at the top of the page. See Figure 3—we provide more detail in §4.3. (AT&T was out of stock on new HTC smart phones; AT&T followed industry practice to focus on refurbished smart phones when new phones were out of stock. Industry experience suggests lower click-through rates for refurbished products, but the decrease should affect the test and control cells equally.)

[Insert Figure 3 about here.]

4.2. CNET Calibration Study

We first identified a candidate set of cognitive-style questions using those suggested by HULB augmented from the references therein and from Novak and Hoffman (2009). We drew 199 consumers from the Greenfield Online panel for a pre-study. These consumers answered all cognitive-style questions. Factor analysis and scale purification identified eleven questions likely to categorize CNET consumers. (Detailed questions and pre-study analyses are available from the authors.)

In the calibration study, 1,292 CNET users answered the eleven purified questions. We

factor analyzed the answers and identified three factors which we labeled impulsive vs. deliberative, analytic vs. holistic, and instinctual vs. not. See Appendix 2. Following standard procedures (e.g., Churchill 1979), we re-purified these scales resulting in three multi-item bipolar cognitive-style dimensions with reliabilities of 0.75, 0.66, and 0.57, respectively. CNET's designers felt they could target most effectively consumer segments that varied on the two most-reliable cognitive-style dimensions. We follow the methods in HULB and assign consumers to segments based on median splits of the two bipolar scales. The four segments were deliberative-holistic, deliberative-analytic, impulsive-holistic, and impulsive-analytic. While the dimensions are orthogonal by construction, there is no reason that the four segments contain equal numbers of consumers. *In vivo* posterior estimates were 9%, 42%, 23%, and 27%, respectively.

4.3. Banner Characteristics (designed by CNET's agency)

CNET's agency varied characteristics of the morphs such as the smartphone image (home screen vs. pictures of people), the size of the image, the size and colors of the fonts, the background colors, and information content (online only, free shipping vs. a list of smart phone features). The designers also varied hot links such as "get it now," "learn more," "watch video," "benefits of Android technology," "see phone details," and "offer details." With a moderate number of banners it was not feasible to vary all of these banner characteristics in a fractional factorial. Rather, we relied on CNET's designers to provide banners that varied substantially. The morphing algorithm automatically and optimally assigned the banners to latent segments (via the Gittins' indices). CNET's choice of potential banners is an empirical tradeoff—more banners might achieve greater discrimination, but more banners might compromise the optimal policy by spreading updating over a greater number of morph x segment indices. More empirical experience might suggest procedures to determine the number of banners that optimizes this

tradeoff.

CNET's agency relied on the judgment of their designers. With more empirical experience, banner designers will be better able to design banners to target latent segments. Researchers might use pre-studies to link banner characteristics to segments identified in the calibration study. Analyses similar to the logit model that links click preferences to segments could help designers select banner characteristics.

4.4. Calibrated Model of Segment-Specific Click Preferences

We observed the clickstreams for all 1,292 consumers in the calibration study. We decompose every click alternative into a vector of 22 click characteristics including dummy variables for areas on the homepage ("carousal," "navigation bar," "promotion bar," "more stories," "popular topics," etc.), areas on other pages (product-specific reviews, "CNET says," "inside CNET," etc.), usage patterns (search category, social influences, tech-savvy news, etc.), and independent judges' evaluations of expected click outcomes (pictures, graphs, data, etc.). The same decomposition applied to the website in the calibration study and to the tracked areas of the *in vivo* website. Using the calibration data we estimated segment-specific click-characteristic weights, $\vec{\omega}_r$. The specific model mapping characteristics to clicks is a logit model that is conditioned upon segment membership. See Appendix 1 (Equation A1) and HULB (p. 211, Equation 4). Parameter values are given in Appendix 3.

4.5. Posterior Beliefs about Latent Cognitive-Style Segments *in vivo*

During day-to-day operation on the CNET website, we use Bayesian updating to estimate the probabilities that each consumer belongs to each latent segment. See Appendix 1 (Equation A1) and HULB (p. 211, Equation 5). Simulations based on the calibration study suggested that five clicks ($t_0 = 5$) would provide sufficient observations to obtain reasonable posterior esti-

mates of $\Pr(r_n = r | \vec{c}_{nt})$.

In CNET, unlike in HULB, we use cookies so that updating continues through multiple consumer visits to CNET. We define an active consumer as a consumer who has made at least five clicks on tracked areas of the website. In the control cell we track clicks but only to determine whether a consumer is active. Before becoming active, consumers are not shown any banners (neither in test nor control). After becoming active, test consumers see a banner selected by the morphing algorithm and control consumers see a randomly chosen banner.

4.6. Defining a Successful Click-through when there are Multiple Sessions

The same banner might be shown in many sessions. (CNET considers a session new after 30 minutes of inactivity.) CNET (and AT&T) consider the banner a success if the consumer clicks through in at least one session. We adopt their definition when we update the p_{rm}'s. To account for interrelated sessions, we use a strategy of temporary updates and potential reversals.

This strategy is best illustrated with a three-session example. Suppose that a consumer sees the same banner in three sessions and clicks through only in the second session. A naïve application of HULB would make three updates to the parameters of the posterior distributions for the success probabilities, p_{rm}. The updates would be based erroneously on observations classified as a failure, then a success, and then a failure. Instead, using CNET's success criterion, the correct posterior is computed after the third session based on one success because the banners achieved their collective goal of at least one consumer click-through. Until we reach the third session, updates should represent all information collected to that point. We update as follows. After the first session (no click through), we update the posterior distribution based on a failure—this is the best information we have at the time. After the second session (click through), we reverse the failure update and update as if success. After the third session (no click through),

we do nothing because the update already reflects a success on CNET's criterion. The mathematical formulae for CNET's success criterion are given in Appendix 1.

4.7. Priors for Morph x Segment Probabilities (as Used in Computing Indices)

The morphing algorithm requires that we set priors for the morph x segment click-through probabilities. The findings in HULB suggest that weakly-informative priors suffice. We set priors equal to the historic click-through probability for banners for refurbished smart phones—the same for all banners. To ensure that the priors are weakly informative, we select parameters of the prior distribution based on an effective sample size of forty consumers—small compared to the anticipated number of CNET consumers.

4.8. Interaction between Morphing and Context Matching

CNET uses context matching, thus one goal of the field experiment was to determine whether morphing adds incremental lift. The context-matching literature reports lifts of approximately 3% for *in vivo* testing and 26% for in-sample projections (see §2). These lifts were calculated for banners or page-views (not on a consumer-by-consumer basis).

The information technology literature consistently postulates that context matching is effective because the banner is more relevant to the consumer (e.g., Chen, Pavlov, and Canny 2009, p. 209; Chu, et al. 2009, p. 1103; Joshi, Bagherjeiran, and Ratnaparkhi 2011; p. 59). Relevance has a long history in advertising research. For example, classic studies postulate that "persuasion may work best depending on whether … message-relevant thought occurs" (Petty, Cacioppo, and Schumann 1983, p. 137). Chaiken (1980, p. 760) manipulates issue involvement as "personal relevance" and demonstrates that greater quality advertising is more persuasive with high involvement but not with low involvement. Zaichkowsky (1986, p. 12) summarizes that "although there does not seem to be a single precise definition of involvement, there is an under-

lying theme focusing on personal relevance." Her survey of the literature indicates that "under high involvement, attitudes were influenced by the quality of the arguments in the message." Prescriptive theories of targeting make similar predictions (Iyer, Soberman and Villas-Boas 2005; Kenny and Marshall 2000).

If these theories apply to banner advertising, and if morphing increases the effective quality of the communication, then we expect an interaction between morphing (increased quality) and context matching (relevance). If cognitive-style matching makes it easier for consumers to learn their preferences, then a morphing-by-context-matching interaction is also consistent with observed interactions between targeting and preference learning (Lambrecht and Tucker 2011).

In our field experiment we manipulate morphing (test vs. control) randomly. Within each experimental cell, some banners match context and some do not. Context-matching occurs naturally on the CNET website and occurs in the same manner in the test cell as in the control cell.

4.9. Results of the CNET Field Experiment

CNET placed banners on their website for all active consumers in the test and control cells during April 11, 2011 to May 13, 2011. Naturally, there were non-AT&T-HTC banners placed on CNET during the 31-day test period, but these banners were placed randomly between test and control. Both we and CNET went to great lengths to ensure there were no systematic effects of non-AT&T-HTC banners or interactions with AT&T-HTC advertising. Sampling appeared random—we detected no systematic differences in the placement of control banners across estimated (latent) cognitive-style segments ($\chi^2_{30} = 15.9, p = 0.98$).

Table 1 summarizes the field-test results. Overall, 116,168 consumers saw 451,524 banners. Of these, 32,084 consumers (27.4%) saw 58,899 banners (13.0%) on webpages where any smart phone was rated, compared, priced, discussed, or pictured. We consider such webpages as

context-matched. Consistent with theories of relevance-quality interactions, morphing achieves significant and substantial incremental improvements for banners on context-matched webpages ($t = 3.0, p = 0.003$). Because many consumers saw multiple banners, we also calculate click-through rates on a consumer-by-consumer basis. Morphing is significantly better than the control on consumer click-through rates when the banners are placed on context-matched webpages ($t = 2.2, p = 0.028$).

Morphing almost doubled click-through rates for context-matched banners (83% and 97% lifts, respectively for banners and for consumers). To put these lifts in perspective, context-matching alone achieved a 5% lift in banner click-through rates, but the difference was not significant ($t = 0.3, p = 0.803$). A 5% lift is consistent with Joshi, et al. (2011) and Chu, et al. (2009) who report lifts of 3.3% and 3.2%, respectively, on large samples. Context-matching alone had a negative lift on consumer click-through rates, but the lift was not significant ($t = 1.4, p = 0.167$).

[Insert Table 1 about here.]

Table 1 also suggests that gains to morphing require that the banners be relevant to the webpage visited by the consumer. There was a decline for banners and consumers when the banners were not on context-matched webpages ($t = 0.5, p = 0.495$ and $t = 1.74, p = 0.081$, respectively), but that decline is marginally significant at best. Interactions between morphing and context-matching were significant for banners ($\chi^2 = 161.8, p < 0.01$) and for consumers ($\chi^2 = 8.2, p = 0.017$).

4.10. Morphing Discriminates Among Latent Cognitive-Style Segments

Segment membership is latent; we do not observe segment membership directly. Instead we use posterior estimates of segment membership to examine the probability that morph m was

assigned to segment r. Table 2 reports posterior probabilities for square banners and for top-of-page banners. On average, top-of-page banners did better than square banners—a result that does not appear connected to cognitive-style morphing. For example, in the context of search advertising, eye-tracking studies suggest a "golden triangle" or "F-shaped" attention pattern; top-of-page sponsored links receive substantially more attention than right-of-page sponsored links (Buscher, Dumais, and Cutrell 2010). Buscher, et al. (p. 47) suggest further that high-quality sponsored links receive twice as much visual attention as low-quality sponsored links. For ease of comparison between different types of banners we renormalize click-through rates for top-of-page banners.

Morph-to-segment matching worked well for square banners. Some square banners are differentially better for specific cognitive-style segments. For example, the best morph for the deliberative-analytic latent segment ($r = 2$) is Morph S1. The best morph for the impulsive-holistic segment ($r = 3$) is Morph S5. The morphing algorithm discriminates less well for the deliberative-holistic segment ($r = 1$), likely because that segment is a much smaller segment than the others (9% of the consumers).

[Insert Table 2 about here.]

The deliberative-analytic segment ($r = 2$) and the impulsive-holistic segment ($r = 3$) together account for 65% of the consumers and each received their best morphs most often. The morphing algorithm does less well for the remaining 35% of the consumers. For the impulsive-analytic ($r = 4$) segment and the deliberative-holistic segment ($r = 1$), the best morph was given more often than average, but other morphs were given more often.

The posterior probabilities for top-of-page banners illustrate a situation where designers did not achieve enough variation. The algorithm learned correctly that Morph T1 was best for all

latent segments. Overall, the morph assignments were enough to achieve substantial lift, but the lift would likely have improved if the algorithm had run longer. When click-through rates are low, the CNET data suggest convergence even beyond 82,000 consumers. This result illustrates why large samples are necessary to evaluate *in vivo* banner morphing.

We attempted to link the features of the best morphs to cognitive style segments. Some assignments made sense. For example, the best morph for the deliberative-analytic segment included a detailed list of product features and the best morph for the impulsive-holistic segment included a link to "get it now." We are hesitant to over-interpret these qualitative insights because, in the CNET field test, there are many more features than morphs.

5. Automotive Experiment to Test Matching Morphs to Segments

Banner advertising generates click-throughs, but banners are also display advertising and may enhance a brand's image whether or not a consumer clicks through. For example, Nielsen (2011) describes a survey in which "54 of those surveyed believe online ads are highly effective at 'enhancing brand/product image.'" Because managers are often interested in more than click-through rates, we supplement the CNET field experiment with an *in vitro* automotive experiment. (Organizational differences between CNET and AT&T, and proprietary concerns, made it impossible to track click-through rates back to sales of AT&T telephones.) In the automotive experiment we abstract from the mechanics of Gittins' learning to test whether morph-to-segment matching increases brand consideration and purchase likelihood as well as click-through rates. The automotive experiment enables us to further test the hypothesis that banner advertisements are more effective when targeted to consumer segments that vary on cognitive styles.

Measures of brand consideration and purchase likelihood require intrusive questions, unlike measures of click-through rates which can be observed unobtrusively. To measure brand

consideration and purchase likelihood we invited consumers to complete questionnaires before and after searching for information on an automotive information-and-review website. Because a sample size of tens of thousands of consumers was not feasible with this design, we used longitudinal methods as a surrogate for dynamic program optimization. Figure 4 summarizes the longitudinal methods. In Phase 1, consumers rated all test and control advertisements for their buying stage and preferred body-type. Two weeks later, in Phase 2, consumers answered a series of questions that enabled us to assign consumers to cognitive-style segments. In Phase 2 we also obtained pre-measures of brand consideration and purchase likelihood. Phases 1 and 2 replaced Bayesian inference and Gittins'-index-based optimization with *in vitro* measurement. Phases 1 and 2 assigned each consumer to a segment and identified the best banners for each segment, thus replacing two tasks performed *in vivo* in the CNET experiment. The actual experiment, Phase 3, occurred two and one-half weeks after Phase 2 (4½ weeks after consumers rated banners in Phase 1). In the experiment (Phase 3) consumers saw banners while exploring an automotive information-and-review website. In the test cell, banners were matched to cognitive styles (plus buying stage and body-type preference) while in the control cell banners were matched only to body-type preference. (Note that this experiment extends the definition of consumer segments to include buying stage—a practical consideration in the automotive market.)

[Insert Figure 4 about here.]

The experimental design, its implications, and potential threats to validity are best understood and evaluated within context. Thus, before we describe the Phase 3 experiment, we first describe the website, the automotive consumer segments, and the test and control banner advertisements.

5.1. Automotive Banners on an Information-and-Recommendation Website

Information-and-recommendation websites, such as Edmunds', Kelley Blue Book, Cars.com, and AutoTrader, play major roles in automotive purchasing. For example, Urban and Hauser (2004) estimate that at least 62% of automotive buyers search online before buying a car or truck. More recently, Giffin and Richards (2011) estimate that 71% of automotive buyers search online and that online search was more influential in purchase decisions than referrals from family or friends, newspapers, and other media sources. Because information-and-recommendation websites attract potential purchasers, automotive manufacturers invest heavily in banner advertising on these websites. The importance of such expenditures motivated General Motors to test morph-to-segment-matching for banner advertising targeted for their Chevrolet brand. General Motors' managerial motivation matched our scientific desire to test whether morph-to-segment matching would enhance brand consideration and purchase likelihood.

We created a website that simulated actual information-and-recommendation websites. Figure 5 illustrates the landing page and an example search page. Consumers could search for information, receive tips and reviews, learn about insurance, and read reviews just like they would on commercial information-and-recommendation websites. To mimic best practices, all test and control banners were targeted by consumers' expressed preferences for one of five body types. Such targeting is typical on commercial websites. For example, Edmunds.com displays body-type category links (coupe, convertible, sedan, SUV, etc.) prominently on the landing page and uses click-through information from these links to place relevant banner advertising on subsequent webpages and site visits. Body-type targeting enhances external validity and relevance. (Recall that morphing was most effective on relevant CNET webpages.)

[Insert Figure 5 about here.]

21

5.2. Cognitive Styles and Stage of the Automotive Buying Process

Body-type preference and the automotive buying stage were measured in Phase 1; cognitive styles were measured in Phase 2. General Motors defines buying-stage segments by: collection, comparison, or commitment. "Collection" segments included consumers who indicated they were more than a year away from buying a car or truck, but in the process of collecting information. "Comparison" segments included consumers less than a year away from buying a car or truck and who had already gathered information on specific vehicles or visited a dealer. "Commitment" segments included consumers who plan to purchase in the next three months, who have collected information on specific vehicles, and who have visited a dealer.

To identify cognitive styles we asked consumers in a pre-study to answer twenty-nine questions adapted from HULB and Novak and Hoffman (2009). We factor analyzed their answers to identify three factors. Based on the questions that load together, we labeled the first two factors as rational-vs.-intuitive and impulsive-vs.-deliberative. The third factor was hard to define. See Appendix 2. Following standard procedures (e.g., Churchill 1979), we purified the scales resulting in three multi-item cognitive-style dimensions with reliabilities of 0.87, 0.87, and 0.36, respectively. Because morphing requires a moderate number of discrete segments, we defined four cognitive-style segments by mean splits on the first two cognitive dimensions.[3,4] The four segments were rational-impulsive, rational-deliberative, intuitive-impulsive, and intuitive-deliberative.

---

[3] Despite differences in the underlying questions, the type of consumer, and the buying context, the cognitive dimensions for high-tech consumers and automotive consumers were not dissimilar. For each set of consumers, one dimension was impulsive vs. deliberative. The other dimension was either analytic vs. holistic (high tech) or rational vs. intuitive (automotive). More experience might identify common dimensions that can be used across applications. If cognitive dimensions are situation-specific, then research might identify a paradigm that relates the dimensions to a set of defined situations.

[4] In the automotive experiment GM used mean-splits rather than median-splits to define segments. There is no reason to believe this will affect the results. Indeed, the two categorizations are quite similar. When we correct for the differences between median- and mean- splits, the test group is still significantly better than the control group.

5.3. Test and Control Banner Advertisements

Banner designers created test banners that varied on characteristics they judged would appeal to consumer segments with different cognitive styles. Some banners emphasize information; others compare targeted vehicles to competitors, and still others stress test drives, finding a dealer, and purchase details. The banners also varied on the size of the images, the number of images, the amount of information provided, the size of the headlines, the amount of content in the headlines, whether content emphasized product features or recommendations, and other design characteristics. Clicks on banners took consumers to different target webpages (as promised in the banners). The designers judged that these characteristics provided sufficient variation for Phases 1 and 2 to target the banners to each cognitive-style segment. In total there were 75 test banners: (five variations to appeal to different cognitive styles) x (three variations to appeal to different stages of the buying process) x (five variations using Chevrolet vehicles chosen to appeal to consumers interested in different body types). Figure 6 provides examples of 15 test banners for one body type (Chevrolet Tahoe).

[Insert Figure 6 about here.]

In Phase 1, consumers evaluated potential test (and control) banners on meaningfulness, relevance, information content, and believability. Using the average score on these measures we identified the best two test banners for each consumer segment. In Phase 3, consumers in the test cell saw the banners that were matched to their segment. Consumers in the control cell saw the control banners. We allowed consumers' preferences to override designers' prior beliefs just as in the CNET field experiment the dynamic program overrode designers' prior beliefs.

There were ten control banners: two banners for each of five body types. Control banners did not vary by cognitive style or buying stage. The control banners were the banners that Chevrolet was using on real information-and-recommendation websites at the time of the automotive

23

experiment.

The control banners in Figure 6 were most relevant to General Motors' business deci-sions, but if we are to use them as a scientific control we must establish they are a valid control. The literature uses a random selection of "morphs" as a no-morphing control. If General Motors' current banners are better than a random selection of test banners, then any differences between test and control cells would underestimate the gain due to morph-to-segment matching. We could then conclude that the improvement due to matching is at least as large as we measure. However, if current banners are worse than a random selection of test banners, then we could not rule out that the test banners are, on average, simply better than the control banners.

The average score for a test banner is 3.36 (out of 5); the average score for a control ban-ner is 3.70. The combined control banners have significantly larger average scores than random test banners ($t = 10.3, p < 0.01$). For a stronger comparison we compare the two best test ban-ners to the two control banners. Even in this comparison the average test score is still less than the control score ($t = 2.7, p < 0.01$). We therefore conclude that the current Chevrolet banners are a sufficient control. If morph-to-segment matching is superior to the current Chevrolet ban-ners, then it is highly likely that morph-to-segment matching will be superior to either a random-ly-selected set of test banners or to a non-matched mix of the two best test banners.

5.4. Experimental Design and Dependent Measures

In Phase 3 consumers were invited to explore an information-and-recommendation web-site called "Consumer Research Power." Consumers searched naturally as if they were gathering information for a potential automotive purchase. They did so for a minimum of five minutes. While consumers searched, we recorded click-throughs on the banners. During this search, we placed banner advertisements for Chevrolet models as they would be placed in a natural setting.

Test consumers received banners that alternated between the best and second-best banner for their cognitive-style and buying-process segment. Control consumers received banners that alternated between the two control Chevrolet banners.[5] All banners, both test and control, were targeted by body-type preference.

Consumers who clicked through on banners were redirected to various websites—websites that varied by banner (and hence consumer segment). For example, banners targeted to impulsive consumers in the commitment buying stage linked to maps of nearby dealerships while banners targeted to rational consumers in the commitment buying stage linked to information on loans, purchasing, and options packages. We balanced the variety of click-through targets to include enough variation to implement targeting by segment, but not so much that consumers were directed outside the *in vitro* web environment. Our *in vitro* targeting likely underestimates variation obtainable *in vivo* and is, thus, conservative.

After consumers completed their search on "Consumer Research Power," we measured Chevrolet brand consideration and purchase likelihood (post measures).

5.5. Potential Threats to Validity

One potential threat to validity is that exposure to banners in Phase 1 might have contaminated the Phase 3 measures. We took steps to minimize this threat. The Phase 1 questionnaire was relatively short (five minutes) and occurred 4½ weeks before the Phase 3 experiment. In Phase 1 consumers were not allowed to click through on the banners and, hence, did not receive the same rich information experience as in Phase 3. Instructions were written carefully to disguise the goals of the later phases—consumers believed the Phase 3 website experience was a

---

[5] Control consumers also received a more-general banner on the landing page. This more-general banner mimics *in vivo* practice. When we include the more-general banner in our analyses, the exposure-weighted rating of all control banners (3.75) remains significantly better than the exposure-weighted rating of the test banners (3.46), reaffirming the control as a valid control ($t = 3.0, p < 0.01$). To be conservative, we do not include clicks from landing-page banners for either the test or control cells.

test of the website, not an advertising test. We believe that the time delay, the number of banners rated, the lack of active click-through in Phase 1, and instructions that disguised later phases combine to limit contamination from Phase 1 to Phase 3.

More importantly, the experimental design minimizes potential false positives that might be due to contamination. First, Phase 2 is more proximate in time than Phase 3. Contamination, if any, should be larger in Phase 2 than in Phase 3, making it more difficult to show an effect on Phase-3-vs.-Phase-2 measures. Second, contamination, if any, would affect test and control cells equally and have no impact on statistical tests of differences that are invariant with respect to constant effects.

Another potential threat to validity is that the morph-to-segment test chooses from more banners than the control. If a consumer saw a greater variety of banners in the test cell, then we would be concerned about biases due to wear-out in the control cell or biases because of greater variety in the test cell. All else equal, greater variety in the banners that a consumer actually sees increases the odds that a banner is the best banner for a consumer. Our design minimizes this threat because consumers in both test and control cells saw only two different banners.

5.6. Results of the Automotive Experiment Testing the Behavioral Premise of Morphing

We invited 2,292 members of the Gongos Automotive Panel to participate in a multi-phase study of website design. Consumers were screened so that they were an equal or sole decision maker in automotive purchases and planned to purchase a new car or truck in less than three years. This mimics standard practice. Of these, 1,299 consumers agreed to participate (61% response rate) and 588 consumers completed Phases 1, 2 and 3 (45.3% completion rate). More consumers were assigned to the test cell (70%) than the control cell (30%) so that we had sufficiently many consumers in each consumer segment. All statistical tests take unequal cell sizes in-

to account. Dependent measures included click-through rates for banners, click-through rates per consumer, post-measures of brand consideration and purchase likelihood, and comparisons of brand consideration and purchase likelihood between the post-measures (after Phase 3) and the pre-measures (during Phase 2).

5.7. Test-vs.-Control Analyses (Post Only)

Because the pre-conditions were the same in the test and control cells, we begin with post-only results. Table 3 reports the post-only comparisons for the morph-to-segment-matching experiment. As in the CNET field experiment, on body-type-relevant webpages, the lift in click-through rates is significant. The test-vs.-control difference in click-through rates is significant whether we focus on impressions (245% lift, $t = 3.3, p < 0.01$) or consumers (66% lift, $t = 4.4, p < 0.01$). The automotive experiment enables us to look beyond click-through rates to brand consideration and purchase likelihood. Both measures increase significantly based on morph-to-segment matching with consideration the most substantial (30% lift, $t = 4.9, p < 0.01$ and 8% lift, $t = 4.1, p < 0.01$, respectively). As a test of face validity, Chevrolet brand consideration is roughly 29% on a nationwide basis—comparable to the 32% measured in the control cell.

[Insert Table 3 about here.]

Table 3 compares all consumers in the test cell to all consumers in the control cell whether or not they clicked on a banner. We gain insight by comparing those consumers who clicked on a banner to those who did not. The comparison of clickers to non-clickers is consistent with self-selection; brand consideration is 45% higher ($t = 2.9, p < 0.01$) and purchase likelihood is 14% higher ($t = 13.5, p < 0.01$) for clickers vs. non-clickers.

Brand consideration improved for both non-clickers (22% lift) and clickers (17% lift);

purchase likelihood improved for non-clickers (9% lift) and stayed the same for clickers (0%

lift). Recall that these relative lifts are computed on a higher base for clickers than non-clickers

because both brand consideration and purchase likelihood are substantially higher for clickers.

We consider these results tentative because the test vs. control lifts are not statistically significant

when we split the sample to within clickers or non-clickers. Nonetheless, the results are at least

consistent with a hypothesis that the banners acted as display advertising.

5.8. Test-vs.-Control and Pre-vs.-Post Analyses

We increase statistical power by accounting for the pre-measures (as in differences of dif-

ferences) and for variation in segment membership or demographics due to stochastic variation

in random assignment. Table 4 reports the results where we control for pre-measures, segment

membership, and demographics. Click-through and brand consideration are quantal measures

(click or not; consider or not), therefore we use a logit formulation for these measures. Purchase

likelihood is a scaled measure, so a regression suffices. Click-through (all banners) and brand

consideration are significant at the $p < 0.01$ level and purchase likelihood is significant at the

$p = 0.02$ level. Click-through (per consumer) is marginally significant at the $p = 0.06$ level.

[Insert Table 4 about here.]

In Table 4 we used the pre-measure as an independent variable because the pre-measure

accounts for both measurement error and, partially, for unobserved heterogeneity in consumers'

propensity to consider or purchase Chevrolet. We can also remove unobserved heterogeneity

with double-difference formulations. When we do so, test vs. control is significant at the 0.01

level for both brand consideration and purchase likelihood (details from the authors).

Together Tables 3 and 4 suggest that morph-to-segment matching increases brand con-

sideration and purchase likelihood (for automotive consumers) as well as click-through rates. In

addition, morph-to-segment matching may have improved overall brand image even among consumers who did not click through. When combined with the CNET field experiment, the automotive experiment suggests that the effectiveness of banners improves when morphing targets banners to consumer segments.

6. Implications and Future Directions

Online morphing is a nascent technology for improving the effectiveness of banner advertising. HULB established the potential for increasing sales if websites morphed their look and feel, but the evaluation was based on data generated in a calibration study. Subsequently, Hauser, Urban, and Liberali (2012) demonstrated that website morphing could be implemented *in vivo*, but their sample size was not sufficient to establish that the improved outcomes were significant.

The CNET field experiment establishes that an Expected Gittins' Index policy enables an *in vivo* website to learn automatically the best morph for each consumer segment. Click-through rates improve substantially for context-matched (relevant) webpages on a high-traffic website. The automotive experiment establishes that morph-to-segment matching also increases brand consideration and purchase likelihood.

The Expected Gittins' Index provides near optimal learning; we know of no better strategy. By the principle of optimality, the Expected Gittins' Index policy is superior to a policy that sets aside the first N_{large} consumers for a random-assignment experiment. On high traffic websites with low click-through rates, the improvement over an N_{large} policy can be substantial.

6.1. Strategic Implications

When morphing increases click-through rates, the marginal return to banners increases. As firms re-optimize their advertising spending they will allocate proportionally more to banners and less to more traditional media. However, there is a fixed cost to the development of multiple

banners for use in morphing. The targeted banners for the automotive experiment would have cost $250,000 to produce if done at market rates (private communication). For high-volume brands, as in our tests, the incremental improvements in click-through rates, consideration, and purchase intentions justify the fixed cost. For smaller websites or advertisers, the fixed cost may be too steep a price to pay.

The effect of increased banner productivity on total advertising spending is ambiguous and dependent upon the detailed marginal costs and revenues. Addressing this question requires meta-analyses across a variety of product categories, media, and countries. Such meta-analyses are now underway through a consortium of researchers and should provide insights on the future of media spending.

6.2. Norms Rather than Calibration Studies

State-of-the-art morphing technology requires a calibration study to (1) establish the definitions of consumer segments and (2) obtain data on click preferences for each segment (the $\vec{\omega}_r$). We envision future applications that rely on norms rather than calibration studies. For example, in the applications to date the definitions of the cognitive-style segments were somewhat similar. With more applications, we might use meta-analyses to stabilize cognitive-style definitions so that we might identify segments without a calibration study. Researchers might also investigate whether consumers' cognitive styles (as they relate to websites and banners) are inherent characteristics of consumers or whether cognitive styles vary based on the situation the consumer faces as he or she navigates the website.

Similarly, meta analyses might provide strong priors for segment-based click-characteristic preferences, $\vec{\omega}_r$. We might also identify the click-alternative characteristics that best distinguish consumer segments. Such empirical generalizations would enable a website or

an advertiser to rely on norms or an abridged calibration study. A similar diffusion of knowledge

has taken place in pretest market simulators for consumer packaged goods. Initial studies ex-

plored methods, but later studies built the normative databases. Today, most pretest-market fore-

casts rely on norms. When norms become established, we expect morphing to flourish.

6.3. Practical Challenges

The banner-morphing experiments in this paper, and the prior website-morphing tests, re-

lied on experienced designers to develop banners or websites to match consumer segments.

Morphing implementation updates priors about which banners are best for each segment. The re-

sults were sometimes non-intuitive and serendipitous and spurred further creative development.

As we gain more experience, we expect that scientific studies will lead to greater insight into the

challenge of designing *a priori* banners that best target latent segments. Such studies are fertile

grounds for new research. Similarly, if banner characteristics are varied systematically, in a de-

sign in which the contribution of each characteristic can be identified, then *a posteriori* research-

ers might gain further insight into the optimal design of banners.

Another practical challenge is transportable code. All code has been specific to the appli-

cation (and open source). Conjoint analysis, hierarchical Bayes, multinomial logit analyses, and

other marketing science methods diffused widely when generalized software became available.

We hope for the same diffusion with banner and website morphing.

Finally, morphing relies on discrete definitions of segments and morphs. We are aware of

research to define morphs by a factorial design of features and to find the best portfolio of

morphs (e.g., Schwartz 2012; Scott 2010). We are unaware of any research to match morphs to

segments that are based on continuous cognitive-style dimensions, but such research would be

interesting.

References

Barr, Colin. 2008. CBS buying CNET in online push. *Fortune Daily Briefing*, CNNMoney.com. http://web.archive.org/web/20080517215301/http:/dailybriefing.blogs.fortune.cnn.com/2008/05/15/cbs-buying-cnet-in-online-push/. Accessed February 2012.

Buscher, G., S. Dumais, and E. Cutrell. 2010. The good, the bad, and the random: an eye-tracking study of ad quality in web search. *SIGIR '10.* (July 19-23, 2010, Geneva, CH).

Chaiken, S. 1980. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39, (5), 752-766.

Chen, Y., D. Pavlov, and J. F. Canny. 2009. Large-scale behavioral targeting. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY. 209-218.

Chu, W. S-T Park, T. Beaupre, N. Motgi, A. Phadke, S. Chakraborty, and J. Zachariah. 2009. A case study of behavior-driven conjoint analysis on Yahoo! Front Page Today module. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1097-1104.

Churchill, G. A., Jr. 1979. A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16 (February), 64-73.

Dahlen, M. 2001. Banner advertisements through a new lens. *Journal of Advertising Research*, 41 (4), 23-30.

Gatarski, R. 2002. Breed better banners: design automation through on-line interaction. *Journal of Interactive Marketing*, 16 (1), 2-13.

Giffin, B. and J. Richards 2011. The role of the internet in the new and used vehicle purchase

process. Polk View (February), https://www.polk.com/knowledge/polk_views/ role_of_internet_in_new_and_used_vehicle_purchase_process. Accessed February 2012.

Gittins, J. C. 1979. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41 (2), 148–177.

Hauser, J. R., G. L. Urban, and G. Liberali. 2012. When to morph: theory and field test of website morphing 2.0. (Cambridge, MA: MIT Sloan School of Management), August.

Hauser, J. R., G. L. Urban, G. Liberali, and M. Braun. 2009. Website morphing. *Marketing Science*, 28 (2), 202-224.

Iyer, G., D. Soberman, and J. M. Villas-Boas. 2005. The targeting of advertising. *Marketing Science*, 24 (3), 461-476.

Joshi, A., A. Bagherjeiran, and A. Ratnaparkhi. 2011. User demographic and behavioral targeting for content match advertising. *Proceedings of the Fifth International Workshop on Data Mining and Audience Intelligence for Advertising (ADKDD 2011)*, 53-60.

Kenny, D. and J. F. Marshall. 2000. Contextual marketing. *Harvard Business Review*, 78 (6), 119-25.

Krishnamurthy, V. and J. 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.

Lambrecht, A. and C. E. Tucker. 2011. When does retargeting work?: timing information and specificity. London Business School. London UK.

Li, H. and J. L. Bukovac. 1999. Cognitive impact of banner ad characteristics: an experimental study, *Journalism & Mass Communication Quarterly*, 76 (2), 341-53.

Nielsen. 2011. Beyond Clicks and Impressions: Examining the Relationship Between Onlin Ad-

vertising and Brand Building. At www.nielsen.com. Visited January 7, 2013.

Novak, T. P. and D. L. Hoffman. 2009. The fit of thinking style and situation: new measures of situation-specific experiential and rational cognition. *Journal of Consumer Research*, 36 (1), 56-72.

Petty, R. E., J. T. Cacioppo, and D. Schumann. 1983. Central and peripheral routes to advertising effectiveness: the moderating role of involvement. *Journal of Consumer Research*. 10 (September), 135-146.

PricewaterhouseCoopers. 2011. IAB Internet advertising revenue report. http://www.iab.net/media/file/IAB_Full_year_2010_0413_Final.pdf. Accessed February 15, 2011.

Schwartz, E. M. 2012. The attribute-based multi-armed bandit for adaptive marketing experiments. Dissertation Proposal at the Wharton School, University of Pennsylvania, Philadelphia, PA.

Scott, S. L. 2010. A modern Bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry*. 26. 639-658.

Stanaland, A. J. S. and J. Tan. 2010. The impact of surfer/seeker mode on the effectiveness of website characteristics. *International Journal of Advertising*, 29 (4), 569-595.

Sundar, S. S. and S. Kalyanaraman (2004). Arousal, memory, and impression-formation effects of animation speed in web advertising. *Journal of Advertising*, 33 (1), 7-17.

Urban, G. L. and J. R. Hauser (2004). "Listening in" to find and explore new combinations of customer needs. *Journal of Marketing*. 68 (April), 72-87.

Zaichkowsky, J. L. 1986. Conceptualizing involvement. *Journal of Advertising Research*, 15, 2, 4-14.

Table 1
CNET Field Test of Banner Advertisement Morphing

	Sample Size		Click-through Rate [a]			Signif-icance
	Test	Control	Test	Control	Lift	
Context-matched webpages						
All banners	40,993	17,906	0.307 [b]	0.168	+ 83%	0.003
Per consumer	22,376	9,708	0.250 [b]	0.127	+ 97%	0.028
Non-context-matched webpages						
All banners	262,911	129,714	0.151	0.160	- 6%	0.495
Per consumer	59,362	24,722	0.144 [c]	0.197	- 27%	0.081

[a] Click-through rates are given as fractions of a percent, e.g., 0.307 of 1%.

[b] Test cell has a significantly larger click-through rate than control cell at the 0.05 level or better.

[c] Test cell has a marginally significantly smaller click-through rate than the control cell at the 0.10 level.

Table 2
CNET Field Test: Posterior Click-through Rates and Assignment Probabilities

	Square Banners					Top-of-Page Banners		
Posterior Click-through Rates [a, b]	m = S1	m = S2	m = S3	m = S4	m = S5			
Deliberative-holistic segment (r = 1, 9%)	.16%	.12%	***.28%***	.27%	.13%			
Deliberative-analytic segment (r = 2, 42%)	***.47%***	.43%	.30%	.42%	.43%			
Impulsive-holistic segment (r = 3, 23%)	.42%	.25%	.24%	.40%	***.44%***			
Impulsive-analytic segment (r = 4, 27%)	.41%	***.65%***	.33%	.54%	.63%			
Posterior $\Pr(m\|r, square)$ [a, b]	m = S1	m = S2	m = S3	m = S4	m = S5			
Deliberative-holistic segment (r = 1, 9%)	14%	7%	9%	18%	***51%***			
Deliberative-analytic segment (r = 2, 42%)	***41%***	5%	35%	7%	12%			
Impulsive-holistic segment (r = 3, 23%)	15%	5%	21%	22%	***37%***			
Impulsive-analytic segment (r = 4, 27%)	11%	24%	30%	10%	24%			
Posterior Click-through Rates [a, b]						m = T1	m = T2	m = T3
Deliberative-holistic segment (r = 1, 9%)						***.35%***	.26%	.26%
Deliberative-analytic segment (r = 2, 42%)						***.45%***	.35%	.39%
Impulsive-holistic segment (r = 3, 23%)						***.45%***	.32%	.33%
Impulsive-analytic segment (r = 4, 27%)						***.63%***	.48%	.35%
Posterior $\Pr(m\|r, top\ of page)$ [a, b]						m = T1	m = T2	m = T3
Deliberative-holistic segment (r = 1, 9%)						***46%***	13%	41%
Deliberative-analytic segment (r = 2, 42%)						***76%***	12%	12%
Impulsive-holistic segment (r = 3, 23%)						***66%***	21%	13%
Impulsive-analytic segment (r = 4, 27%)						***75%***	16%	9%

[a] Largest values in a column are shown in ***bold italics.*** Rows sum to 100%.   [b] Posterior segment sizes are shown in parentheses (percent of total consumers).

Table 3
Automotive Experiment: Banner Advertisement Morphing (Post-only Results)
(All banners are targeted by body-type preference.)

	Sample Size		Outcome Measure [a]			Signif-icance
	Test	Control	Test	Control	Lift	
Click-through rates						
All banners	6,348	2,643	0.97% [b]	0.26%	+ 245%	< 0.01
Per consumer	421	167	15.9% [b]	9.6%	+ 66%	< 0.01
Brand Consideration	421	167	42.8% [b]	32.9%	+ 30%	< 0.01
Purchase likelihood	421	167	3.28 [b]	3.05	+ 8%	< 0.01

[a] Click-through rates are given as percentages. Consideration is a consider-or-not measure that we report as a percentage of the sample. Purchase likelihood is measured with a five-point scale.

[b] Test cell is significantly larger at the 0.01 level.

Table 4
Automotive Experiment: Banner Advertisement Morphing
Controlling for Pre-measures, Segment Membership, and Demographics

	Clicks All Banners		Clicks per Consumer		Brand Consideration		Purchase Likelihood	
	Coefficient	Significance	Coefficient	Significance	Coefficient	Significance	Coefficient	Significance
Intercept	-7.860 [a]	< 0.01	-3.657 [a]	< 0.01	-4.132 [a]	< 0.01	0.091	0.629
Test vs. control treatment	1.246 [a]	< 0.01	0.564 [b]	0.061	0.759 [a]	< 0.01	0.150 [a]	0.019
Pre-measure	—	—	—	—	3.570 [a]	< 0.01	0.801 [a]	< 0.01
Buying-process dummies								
Collect	0.359	0.458	0.580	0.126	0.777 [a]	0.026	0.347 [a]	< 0.01
Compare	0.763 [a]	< 0.01	1.007 [a]	< 0.01	0.503 [b]	0.065	0.105	0.205
Commit	—	—	—	—	—	—	—	—
Cognitive-style segment dummies								
Rational-impulsive	0.417	0.114	-0.038	0.972	0.389 [a]	0.038	-0.001	0.172
Rational-impulsive	0.094	0.290	0.102	0.792	0.221	0.144	0.082	0.142
Intuitive-impulsive	0.537	0.513	0.407	0.259	0.707 [a]	0.022	0.103 [b]	0.059
Intuitive-deliberative	—	—	—	—	—	—	—	—
Male (vs. Female)	-0.414	0.157	-0.344	0.162	0.285	0.247	0.079	0.300
Age	0.025 [a]	0.033	0.020 [b]	0.055	0.014	0.184	0.006 [b]	0.065
Income	0.000	0.547	-0.002	0.451	0.000	0.476	-0.000	0.767
Log-likelihood ratio	-349.845 [a]		-223.039 [a]		-232.013 [a]		-734.298 [a]	

[a] Significant at the 0.05 level.   [b] Significant at the 0.10 level. Sample size is 8.991 impressions or 588 consumers. All equations are significant at the 0.01 level. Test vs. control is also significant at the 0.01 level with a differences of differences specification (available from the authors).

Figure 1
Conceptual Diagram of Banner Morphing (Illustrative Values Only)

Gittins' Indices by
Segment x Morph
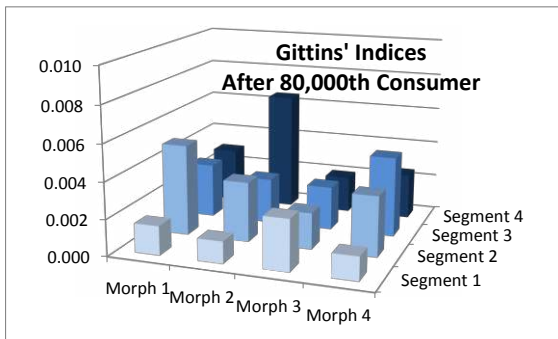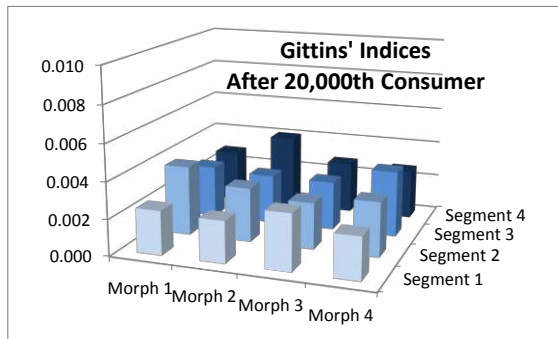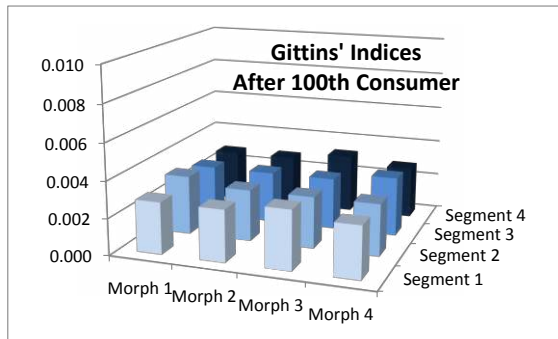
Posterior Probability
that Current Consumer
Belongs to Latent Segment r
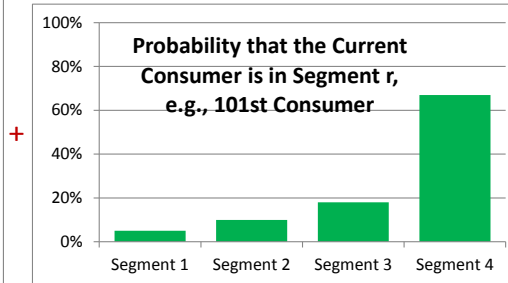
Expected Gittins'
Index by Morph
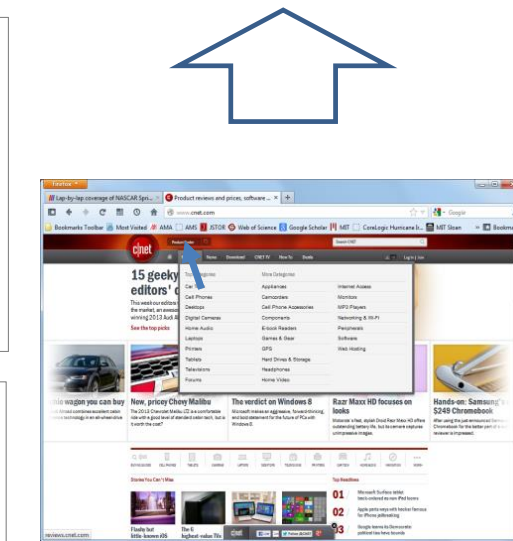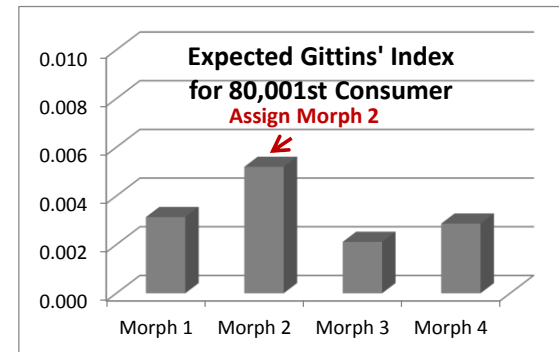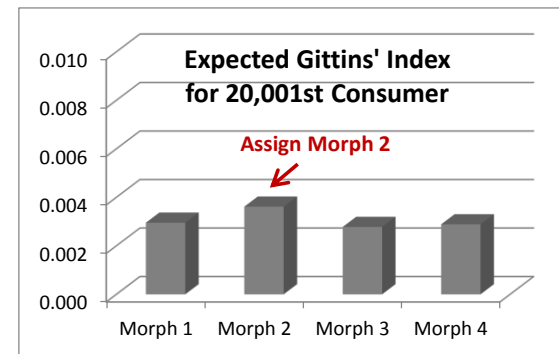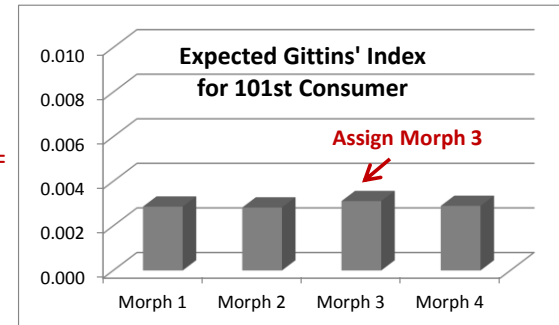
Observe clicks *in vivo* to obtain posterior
probabilities of segment membership.

Figure 2
The Different Roles of the Calibration Study and the Day-to-Day Banner-Morphing Algorithm

Tasks   # Outcomes

Calibration Study
(prior to *in vivo* morphing)

1. Measure cognitive styles with established questions and define cognitive-style segments.

2. Observe clicks and characteristics of clicks for consumers in each cognitive-style segment.

1. Assign each calibration-study consumer to a cognitive-style segment (using questions only in the calibration study).

2. Calibrated model which can infer segment membership probabilities from clickstream.

Day-to-day operation
(of *in vivo* website)

Exploration

Exploitation

1. Observe clickstream. Use calibrated model to infer consumers' latent cognitive-style segments.

2. Observe outcomes (e.g., click-throughs). Update Gittins' indices for each segment x morph combination.

3. Use latent segment probabilities and Gittins' indices to compute Expected Gittins' index.

1. Cognitive-style probabilities for each latent segment.

2. Gittins' index value for each segment x morph combination after the nth consumer.

3. (Near) optimal assignment of a morph to the nth consumer to balance exploration and exploitation.

Figure 3
Square and Top-of-Page Banner Advertisements (CNET Field Experiment)

Figure 4
Automotive Experiment: Longitudinal Design as Surrogate for Morph-to-Segment Matching
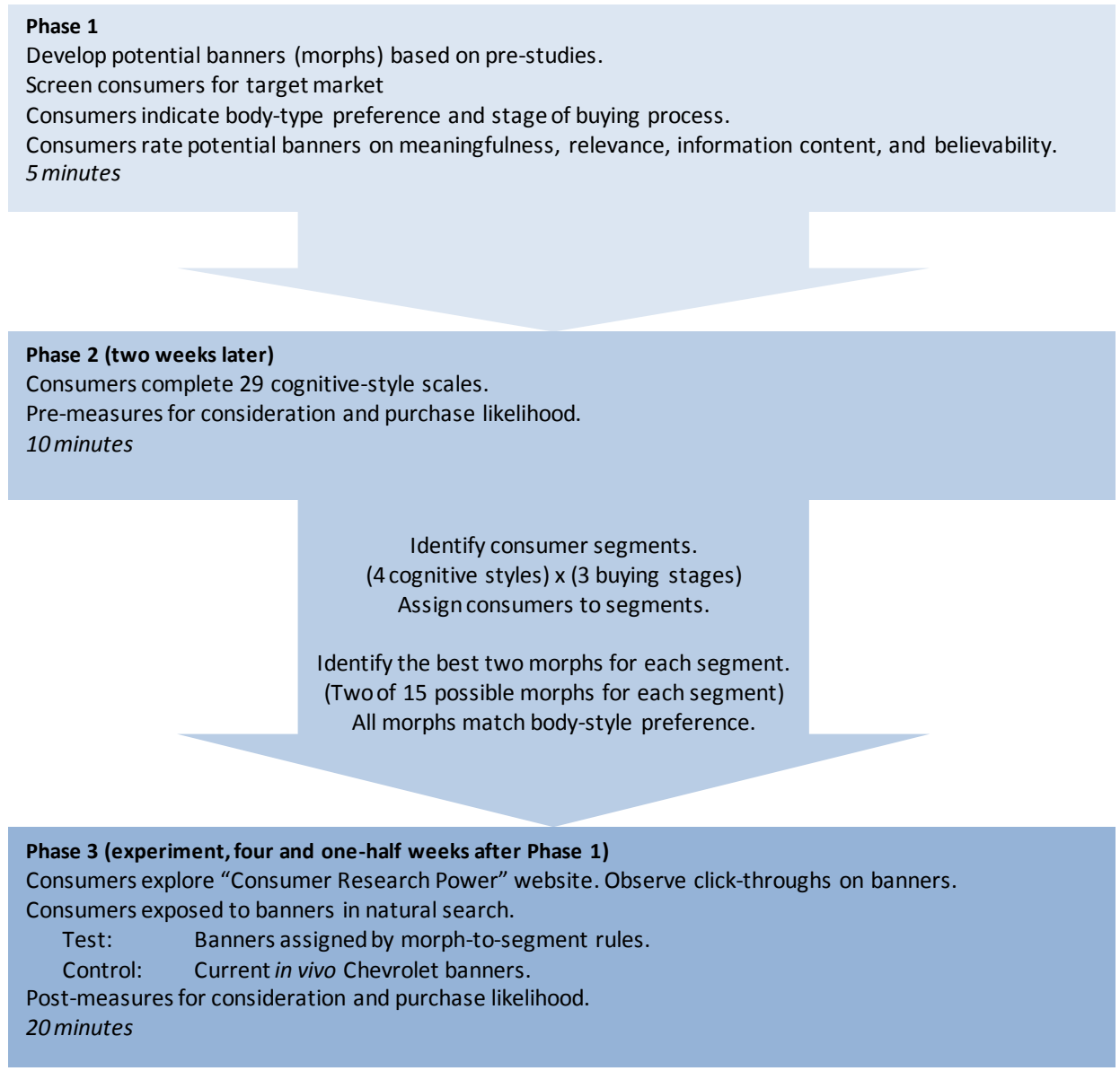*(Phases 1 and 2 replace in vivo Bayesian inference and Expected Gittins' Index optimization.)*

Phase 1
Develop potential banners (morphs) based on pre-studies.
Screen consumers for target market
Consumers indicate body-type preference and stage of buying process.
Consumers rate potential banners on meaningfulness, relevance, information content, and believability.
5 minutes

Phase 2 (two weeks later)
Consumers complete 29 cognitive-style scales.
Pre-measures for consideration and purchase likelihood.
10 minutes

Identify consumer segments.
(4 cognitive styles) x (3 buying stages)
Assign consumers to segments.

Identify the best two morphs for each segment.
(Two of 15 possible morphs for each segment)
All morphs match body-style preference.

Phase 3 (experiment, four and one-half weeks after Phase 1)
Consumers explore "Consumer Research Power" website. Observe click-throughs on banners.
Consumers exposed to banners in natural search.
    Test: Banners assigned by morph-to-segment rules.
    Control: Current *in vivo* Chevrolet banners.
Post-measures for consideration and purchase likelihood.
20 minutes

Figure 5
Simulated Website for Automotive Experiment Matching Morphs to Segments
(Landing page on the left. One of many subsequent pages on the right.)

Figure 6
Example Test and Control Banner Advertisements for the Automotive Experiment
(The left-most banners are controls. The other columns contain five banners designed for each buying-stage segment. In the experiment there were 10 potential control banners: body type x two banners. There were 75 potential test banners: body type x buying-stage x cognitive-style.)

Control	Collection	Comparison	Commitment

Appendix 1. Mathematical Summary of Morphing Algorithm

A1.1. Notation

Let n index consumers, r index consumer segments, m index morphs, t index clicks, and j index click alternatives. Capital letters indicate totals. Let $c_{ntj} = 1$ if n chooses the jth click alternative (link) on the tth click and $c_{ntj} = 0$ otherwise. Let $\delta_{mn} = 1$ if we observe a positive outcome when n sees morph m, and $\delta_{mn} = 0$ otherwise. Let $\vec{c}_{nT}$ be the vector of the c_{ntj} up to and including T^{th} click, let $\vec{x}_{jtn}$ be the vector of characteristics of the jth click alternative for the tth click for consumer n, let $\vec{\omega}_r$ be the vector of preference weights for the $\vec{x}_{jtn}$ for the rth segment, let $\text{Pr}_0(r_n = r)$ be the prior probability that n is in segment r, let q_{rn} be the probability that n belongs to segment r, let p_{rm} be the probability of observing an outcome (sale, click-through, etc.) if a consumer in segment r sees morph m, let G_{rm} be Gittins' index for r and m, and let a be the consumer-to-consumer discount rate.

A1.2. Assigning Consumers to Segments

We first estimate the $\vec{\omega}_r$ from a calibration study in which consumers answer questions to identify their segments. We also we observe the click alternatives they choose. The estimation is based on a logit likelihood using either maximum-likelihood or Bayesian methods. Details are standard, available in HULB, and not repeated here. For online morphing we know the $\vec{x}_{jtn}$'s for key click alternatives. We compute $\vec{x}'_{jtn}\vec{\omega}_r$, which is n's observed utility for the jth click alternative for the tth click. Using the logit likelihood (HULB, p. 211, Equations 4 & 5), we obtain the probability that observed clicks are chosen given that the consumer is in segment r. Bayes Theorem provides q_{rn}.

$$\Pr(\vec{c}_{nT}|\vec{\omega}_r, \vec{x}_{jkn}s) = \Pr(\vec{c}_{nT}|r_n = r) = \prod_{t=1}^{T}\prod_{j=1}^{J_k}\left(\frac{\exp[\vec{x}'_{jtn}\vec{\omega}_r]}{\sum_{\ell=1}^{J_k}\exp[\vec{x}'_{\ell tn}\vec{\omega}_r]}\right)^{c_{ntj}}$$

(A1)

$$q_{rn} = \Pr(r_n = r|\vec{c}_{nT}) = \frac{\Pr\{\vec{c}_{nT}|r_n = r\}\Pr_0(r_n = r)}{\sum_{s=1}^{R}\Pr\{\vec{c}_{nT}|r_n = s\}\Pr_0(r_n = s)}$$

A.1.3. Updating Beliefs about the Probability of an Outcome Given a Morph and Segment

After observing outcomes for each consumer, n, we update our beliefs about outcome probabilities. Call these probabilities p_{rmn}. Using beta-binomial updating we represent posterior knowledge about these probabilities with a beta distribution with parameters α_{rmn} and β_{rmn}. If we knew the consumer's segment with certainty, we could update these parameters with standard formulae. However, segment membership is only partially observable, hence we use the segment-membership probabilities to treat the nth observation as R fractional observations where R is the number of latent segments. The binomial formula is a well-defined probability density function for non-integer values.

$$\alpha_{rmn} = \alpha_{rm,n-1} + \delta_{mn}q_{rn}$$

(A2)

$$\beta_{rmn} = \beta_{rm,n-1} + (1 - \delta_{mn})q_{rn}$$

Equation A2 suffices for website morphing, but for banner morphing the relevant criterion is at least one click-through per consumer. For this criterion we take multiple sessions into account. In banner morphing we use Equation A2 at the end of the first session of a new consumer. Subsequently, if the any prior outcome was a success ($\delta_{mn} = 1$), we do nothing. If all prior outcomes were failures ($\delta_{mn} = 0$) and we observe a failure, we do nothing. If all prior outcomes were failures ($\delta_{mn} = 0$) and we now observe a success ($\delta_{mn} = 1$), we reverse the update.

A2

Prior failures did not change the α_{rmn}'s for each r, so we now add q_{rm}. When a failure becomes a success, we undo the update that was added to the β_{rmn}'s for each r. Earlier failures caused us to add q_{rn} for each r to the β_{rmn}'s, hence we now subtract q_{rn} for each r from the β_{rmn}'s.

A.1.4. Calculating the Gittins' Indices for Each Morph and Segment

First assume the consumer's segment is known. Gittins' Index Theorem enables us to decompose a dynamic program over M morphs into M much simpler dynamic programs. The optimal strategy is to choose in each period the morph with the largest index in that period. Gittins' index provides the needed metric for each uncertain morph by comparing it to a fixed option with a probability, G_{rm}, of a positive outcome. Bellman's equation for the morph-and-segment specific dynamic program is given as follows. (Details are in HULB p. 207-208 and Gittins 1979.) In this equation, $R(\alpha_{rmn}, \beta_{rmn}, a)$ is Bellman's value function. We solve this equation for fixed points to table G_{rm} as a function of α_{rmn} and β_{rmn}. (a is fixed.)

$$(A3) \qquad R(\alpha_{rmn}, \beta_{rmn}, a) = \max \left\{ \begin{array}{l} \dfrac{G_{rmn}}{1-a}, \qquad \dfrac{\alpha_{rmn}}{\alpha_{rmn} + \beta_{rmn}}[1 + aR(\alpha_{rmn} + 1, \beta_{rmn}, a] \\[2ex] \qquad + \dfrac{\alpha_{rmn}}{\alpha_{rmn} + \beta_{rmn}} aR(\alpha_{rmn}, \beta_{rmn} + 1, a) \end{array} \right\}$$

A.1.5. Choosing the Morph in Each Period

When consumer segments are latent, we chose the morph in each period that has the highest value for the Expected Gittins' Index, EG_{mn}. Krishnamurthy and Mickova (1999) show that this expected index identifies a (near) optimal policy.

$$(A4) \qquad\qquad\qquad EG_{mn} = \sum_{r=1}^{R} q_{rn} G_{rmn}(\alpha_{rmn}, \beta_{rmn})$$

Appendix 2. Factor Loadings Matrices for CNET and Automotive Experiments

We factor analyze consumers' self-evaluations on cognitive-style items questions using principle component analysis and varimax rotation with Kaiser normalization retaining factors with eigenvalues greater than one. We interpret the factors based on the factor loadings and then use scale purification with Cronbach's alpha to select scale items (Churchill 1979). Segments are based on retained scales (sufficient reliability). In the calibration study consumers are assigned to segments based on median splits (CNET) or mean-splits (automotive) of sum scores.

Although three cognitive-style dimensions were identified initially for both the CNET and automotive experiments, each experiment used only the first two cognitive-style dimensions to define latent segments. In the CNET experiment, CNET's designers judged they could best target the first two dimensions. In the automotive experiment, the third dimension was difficult to interpret and did not have sufficient reliability. For completeness, we report here all three dimensions for both experiments.

A2.1. Cognitive-Style Factor Loadings for CNET Field Experiment

	Impulsive vs. Deliberative	Analytic vs. Holistic	Instinctual vs. Not
I rely on my first impressions.	0.086	0.208	0.654
I am detail oriented and start with the details in order to build a complete picture.	-0.711	-0.066	-0.057
I find that to adopt a careful, analytic approach to making decisions takes too long.	-0.005	0.699	0.166
I go by what feels good to me.	-0.055	0.289	0.680
When making a decision, I take my time and thoroughly consider all relevant factors.	-0.794	-0.098	0.067
I do not like detailed explanations.	0.220	0.570	0.173
I reason things out carefully.	-0.748	-0.139	0.000
Given enough time, I would consider every situation from all angles.	-0.747	-0.034	0.061
I do not tackle tasks systematically.	0.058	0.753	0.047
I use my instincts.	-0.100	-0.033	0.798
I do not approach tasks analytically.	0.108	0.759	0.103

A2.2. Cognitive-Style Factor Loadings for Automotive Three-Phase Experiment

	Rational vs. Intuitive	Impulsive vs. Deliberative	Ignore Images, Focus on Details
I reasoned things out carefully.	0.71	-0.32	0.01
I tackled this task systematically.	0.58	-0.37	0.15
I figured things out logically.	0.64	-0.33	0.18
I approached this task analytically.	0.62	-0.40	0.16
I applied precise rules to deduce the answer.	0.63	-0.18	0.16
I was very aware of my thinking process.	0.62	-0.24	0.04
I used my gut feelings.	0.29	0.72	0.08
I went by what felt good to me.	0.30	0.69	0.13
I relied on my sense of intuition.	0.41	0.67	0.06
I relied on my first impressions.	0.22	0.66	0.14
I used my instincts.	0.30	0.67	0.11
Ideas just popped into my head.	0.30	0.59	0.05
I tried to visualize the images as 3-D shapes.	0.54	0.24	-0.26
I read the text carefully.	0.57	-0.25	-0.13
I skimmed the text.	-0.18	0.23	0.31
I concentrated on the images.	0.48	0.44	-0.34
I ignored the images.	-0.20	-0.15	0.66
I made comparisons of different facts.	0.53	-0.16	-0.09
I made comparisons between different images.	0.47	0.19	-0.27
I did not notice there were video reviews.	-0.22	0.05	0.58
The video reviews were helpful in making my decision.	0.49	0.29	-0.19
I like detailed explanations.	0.53	-0.21	0.02
I enjoy deciphering graphs, charts and diagrams about products and services.	0.56	-0.19	0.12
I prefer planning before acting.	0.49	-0.31	0.06
I'm usually more interested in parts and details than in the whole.	0.31	0.23	0.43
I like to make purchases without thinking too much about the consequences.	0.11	0.47	0.31
I tend to see problems in their entirety.	0.52	-0.18	0.08
I see what I read in mental pictures.	0.55	0.20	-0.13
I am detail oriented and start with the details in order to build a complete picture.	0.60	-0.23	0.17

Appendix 3. Estimation of $\vec{\omega}_r$ for the CNET Experiment

We follow the procedures detailed in HULB to estimate click-characteristics preferences. We use these values to compute the posterior probabilities for latent cognitive-style segments in real-time. Table A3.1 provides maximum likelihood estimates of $\vec{\omega}_r$. This estimation explains 60.5% of the uncertainty (U^2 [pseudo-R^2] of 0.605). The sample size is likely sufficient; U^2 degrades only to 59.4%, 57.4%, and 56.7% if we use 50%, 33%, and 20% of the data, respectively.

Table A3.1. Maximum-Likelihood Estimates of $\vec{\omega}_r$ for CNET Experiment

	Constant	Segment Indicator Variables	
		Impulsive vs. Deliberative	Analytic vs. Holistic
Expect the linked page to have pictures or graphs	0.257 [a]	0.209	-0.292 [a]
Expect the linked page to be focused on a specific question (technical)	-3.947 [a]	-1.120 [a]	1.351 [a]
Expect the linked page to have large amount of data	1.181 [a]	0.095	-0.221
Navigation Bar	6.931 [a]	-2.459 [a]	2.349 [a]
Carousel	3.946 [a]	0.190	0.665 [b]
More Stories	5.208 [a]	1.053 [a]	0.808 [a]
Promotion Bar	5.762 [a]	-1.853	2.630 [b]
Popular Topics	3.818 [a]	1.517 [a]	-0.981 [a]
Tabs	-14.585	1.236	-0.032
Inside CNET	5.036 [a]	2.597 [a]	-0.858 [b]
Search category	3.706 [a]	-2.856 [a]	2.818 [a]
Product-specific reviews	3.741 [a]	-2.299 [a]	2.083 [b]
Social Influences: expert opinion ("CNET says")	3.360 [a]	1.322 [a]	-1.226 [a]
Social Influences: consumer opinion ("what others do")	2.087 [a]	0.768 [a]	-0.237
Tech-savvy	0.263 [b]	0.036	-0.176

[a] Significant at the 0.05 level.　[b] Significant at the 0.10 level.