Morphing Banner Advertising

by

Glen L. Urban
Guilherme Liberali
Erin MacDonald
Robert Bordley
and
John R. Hauser
July 2012

Glen L. Urban is the David Austin Professor of Marketing, MIT Sloan School of Management, Massachusetts Institute of Technology, E52-536, 77 Massachusetts Avenue, Cambridge, MA 02139. 617 253-6615, glurban@mit.edu.

Guilherme (Gui) Liberali is a Visiting Scholar at MIT Sloan School of Management, and Assistant Professor of Marketing, Erasmus School of Economics, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands, liberali@ese.eur.nl.

Erin MacDonald is an Assistant Professor and Mack 2050 Challenge Scholar, Department of Mechanical Engineering, 2020 Black Engineering Building, Iowa State University, Ames, Iowa, 50011, erinmacd@iastate.edu.

Robert Bordley is Fellow at Booz-Allen-Hamilton, 101 W Big Beaver Rd # 505 Troy, MI 48084-5353 (formerly Technical Fellow at General Motors Research), bordley_robert@bah.com.

John R. Hauser is the Kirin Professor of Marketing, MIT Sloan School of Management, Massachusetts Institute of Technology, E52-538, 77 Massachusetts Avenue, Cambridge, MA 02139, (617) 253-6615, hauser@mit.edu.

This research was supported by the MIT Sloan School of Management, the Center for Digital Business at MIT (ebusiness.mit.edu), GM, WPP/Kantar, Google, CNET.com. We gratefully acknowledge the contributions of our industrial collaborators, research assistants, and faculty colleagues: Dorothee Bergin, Angela Chow, Tousanna Durgan, Shirley S. Fung, Will Hansen, Patricia Hawkins, Douglas Hwang, Tom Kelley, Jong-Moon Kim, Clarence Lee, Jimmy Li, Cordelia Link, Ladan Nafissi, Andy Norton, Jonathon Owen, George Pappachen, Chris Perciballi, Joyce Salisbury, Linda Tan, David Vanderveer, and Kevin Wang.
Morphing Banner Advertising

Abstract

Researchers and practitioners devote substantial effort to targeting banner advertisements, but less effort on how to communicate with consumers once targeted. Morphing enables a website to learn (actively and near optimally) which banner advertisements to serve to each cognitive-style segment in order to maximize click-through, brand consideration, and purchase. Cognitive-style segments are identified automatically 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 a focused experiment on an automotive information-and-recommendation website. The focused experiment replaces automated learning with a longitudinal design to test the premise of 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. Together the field and the focused experiments demonstrate that matching cognitive styles provide significant benefits above and beyond more-traditional targeting. Such improved banner effectiveness has strategic implications for allocations among media.

Keywords: On-line advertising, banner advertising, behavioral targeting, website morphing, cognitive styles, field experiments, electronic marketing, dynamic programming, bandit problems, strategic optimization of marketing.
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 cognitive-style segments which, in turn, are inferred from consumers’ clickstream choices. Example cognitive style segments are impulsive vs. deliberative and analytic/visual vs. holistic/verbal. The website automatically determines the best “morph” for a segment by solving a dynamic program that balances exploration (trying new “morphs” to learn their effectiveness) and exploitation (using advertisements based on current expected outcomes). Advertising morphing modifies methods used in website morphing (Hauser, Urban, Liberali and Braun [HULB], 2009), which changes the look and feel of a website based on inferred cognitive styles. Morphing adds a strategic behavioral science component on top of more-common methods such as context matching and other forms of targeting.

HULB projected a 20% improvement in sales for BT Group’s broadband-sales website, but their projections were based on simulated consumers who behaved according to clickstream preferences and morph x segment probabilities estimated in a priming study. HULB did not have the resources to obtain sufficient in vivo sample to field-test website morphing.\(^1\) (By in vivo we refer to websites visited by consumers for information search or purchasing. By in vitro we refer to experiments in which we gather information from a panel of consumers. In vitro experiments attempt to mimic in vivo field experiments, but never perfectly.)

Large test samples are necessary in vivo because online morphing is designed for high-traffic websites with tens of thousands of visitors. The optimal solution to the dynamic program,

\(^1\) Hauser, Urban and Liberali (2012) report a field test of website-morphing, but with a small sample. Their results are suggestive but not significant. The morphing algorithm may not have reached steady-state on their sample.
Morphing Banner Advertising

on which morphing is based, often includes exploration well past the thousandth consumer per segment. (Technically, the need for so many observations is driven by the relative value of sales from \( n + 1 \)st customer as compared to sales to the \( n \)th customer.) With sixteen consumer segments, HULB estimated a need for 40,000 consumers to realize substantial gains from website morphing (HULB, Figure 3, p. 209).

This first field test of banner morphing is promising. We observe an 83-97% lift in click-through rates between test and control cells above and beyond context matching. Click-through rates are one industry standard used to judge banner effectiveness, but managers also seek brand consideration and purchase likelihood.

To examine additional dependent measures, we supplement the large-sample field test with a smaller-sample random-assignment *in vitro* test of banner-to-segment matching on an automotive information-and-review website. We match banners to consumer segments defined by cognitive styles as well as buying-process stages and body-type preference. We avoid the need for extremely large samples with three longitudinal surveys in which the first two surveys measure advertising preference, cognitive styles, and stage of buying process. In a third survey, separated from the pre-measures by four and one-half (4½) weeks, consumers see banner advertising while searching for information on automobiles and trucks. The sample (588 consumers) is sufficient because we substitute direct measurement for Bayesian inference of segments and for optimal assignment by a dynamic program. *In vitro* measurement enables us to examine the fundamental premise that morph-to-segment matching improves brand consideration and purchase likelihood.

2. Banner Advertising – Current Practice

A banner advertisement is a paid advertisement that appears on a webpage. In the last ten
Morphing Banner Advertising

years online advertising revenue has tripled. Banner advertisements account for 24% of online advertising revenue – about $6.2 billion in 2010. Banner advertisements 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 (PricewaterhouseCoopers 2011, Dahlen 2001). Conversion after click-through varies from 1% to 30% with an average of 2 to 4% (Nielsen 2008; Peterson 2009). As a result managerial interest is high for methods that improve banner effectiveness.

Current theory and practice attempts to increase click-through rates with a variety of methods. The simplest method uses experiments on a training sample to find the best combination of banner features. 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 to search 40 binary features to achieve 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, e.g., “divorce” in a banner to “divorce” on the webpage, but often had trouble distinguishing context, e.g., placing a banner for a divorce lawyer on a gossip account of Christina Aguilera’s divorce (Joshi, Bagherjaeiran and Ratnaparkni 2011). Context matching improved when demographics and click logs were used to match relevant [banner] characteristics to consumers demographics, geographic location, and prior consumption. In a related application to Yahoo!’s news articles rather than banners, Chu, et al. (2009) use these methods to increase click-through rates dramatically. Such 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 based on past behavior. “Behavioral targeting leverages historical user behavior to select the most relevant ads to display (Chan, Pavlov and Canny 2009, p. 209).” These researchers use cookie-based observation of 150,000 prior banners, webpages, and queries to identify which consumers are most likely to respond to banners. They report expected lifts of approximately 26%.

Finally, 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, matching is based on cognitive styles rather than context relevance or past behavior. Second, cognitive-style segments are inferred automatically from the clickstream rather than manipulated (as in surfer vs. seeker research). Third, morphing learns (near) optimally by trading off exploration versus exploitation to identify automatically morph-to-segment matches. Banner morphing is a complement to context matching or behavioral targeting. We expect morphing to be more effective on webpages that are relevant to consumers (and add incremental lift above and beyond existing methods such as context matching or behavioral targeting).

3. Brief Review of Banner Morphing

The fundamental behavioral premise of banner morphing is that a banner advertisement is more effective when it is customized for a cognitive-style segment. Given this premise, an online morphing system must accomplish two tasks to succeed. First, it must assign a consumer to a
Morphing Banner Advertising

segment based on observing the consumer’s clickstream. Second, it must learn the best banner(s) to assign to each consumer segment. Figure 1 provides a conceptual diagram. We describe briefly how HULB’s morphing algorithms accomplish these tasks. Readers familiar with HULB may wish to skip to §4.

[Insert Figure 1 about here.]

3.1. Assigning Consumers to Segments based on Clickstream Data

To calibrate the morphing algorithm we begin with a small-sample priming study. For example, HULB asked 835 broadband consumers to complete a survey in which they answered 13 questions about their cognitive styles, questions such as “I prefer to read text rather than listen to a lecture.” HULB factor analyzed answers to the questions to identify four ipsative cognitive-style dimensions. Median splits on the cognitive dimensions identified sixteen (2x2x2x2) cognitive-style segments.

By observing clicks as consumers explore the priming website, we develop a model that enables us to assign consumers to segments based on the clicks that consumers choose. For example, when faced with multiple click choices, we might observe in the priming study that analytic/visual consumers are more likely than other consumers to choose a tool that enables them to compare broadband plans. Because the priming study tells us how cognitive styles affect clickstream choices, it is natural in day-to-day operations to use Bayesian methods to infer cognitive styles from consumers’ clickstream choices.

Following HULB, let \( r \) index consumer segments, \( m \) index alternative banner advertisements (“morphs”), \( n \) index consumers, \( t \) index clicks, and \( j \) index the various places on a webpage where a consumer might click (called click alternatives). Let \( r_n \) be consumer \( n \)'s segment and \( c_{njt} = 1 \) indicate that consumer \( n \) chose click alternative \( j \) on the \( t^{th} \) click. Let \( \tilde{c}_{nt} \) be con-
sumer $n$’s clicks up to and including the $T^{th}$ click. We decompose each click alternative on a webpage into a vector of click characteristics, $\mathbf{x}_{jtn}$. Click characteristics can be dummy variables for areas of webpage (such as a comparison tool), expectations (click is expected to lead to graphics), or other descriptions. To calibrate the model, we observe clicks, click characteristics, and consumers’ segments in the priming study and use the priming data to estimate a vector of weights, $\omega_r$, for a logit likelihood (details in HULB, p. 211). The estimated weights (from the priming study) enable us to assign a utility, $\mathbf{x}_{jtn}'\omega_r$, to each click alternative on the day-to-day website. We thus form the likelihood of observing the clickstream, $\mathbf{c}_{nt}$. With this likelihood, $\Pr(\mathbf{c}_{nt} | r_n = r)$, prior beliefs, $\Pr_0(r_n = r)$, and Bayes’ Theorem, we estimate the probability, $\Pr(r_n = r | \mathbf{c}_{nt})$, that day-to-day consumer $n$ belongs to segment $r$. We summarize the mathematical details and all notation in Appendix 1.

HULB fix the time to morph, $t_o$, exogenously. For website morphing they set $t_o = 10$ clicks. We adopt their strategy of a fixed time to morph. More complex algorithms have been proposed to determine the optimal time to morph (Hauser, Urban and Liberali 2012), but these algorithms were not available at the time of our experiments. Thus our experiments are conservative because morphing would likely do even better with improved algorithms.

### 3.2. Learning the Best Banner Advertisement for a Consumer Segment

Let $p_{rm}$ be the probability of a good outcome (a sale, brand consideration, or a click-through) given a consumer in segment $r$ experienced morph $m$ for all clicks after $t_o$. One suboptimal method to estimate $p_{rm}$ would be to assign morphs randomly to a large number, $N_{\text{large}}$, of consumers and observe outcomes. This strategy, similar to that used by Google’s web optimizer and many behavioral-targeting or context-matching algorithms, is sub-optimal during the calibration period because $N_{\text{large}}$ consumers experience morphs that may not lead to the best outcomes.
Morphing Banner Advertising

The larger we choose \( N_{large} \), the more consumers are given suboptimal banners.

We minimize opportunity loss during day-to-day operations by solving a dynamic program that balances the opportunity loss incurred while exploring new morph-to-segment assignments with the knowledge gained about the \( p_{rm} \)'s. This knowledge enables us to assign morphs more effectively to future consumers. To solve the dynamic program we parameterize the distributions that summarize our uncertainty about the \( p_{rm} \)'s and use observed outcomes to update the posterior distributions. (The updating equations are in Appendix 1).

If we knew a consumer’s segment, the solution to the dynamic program would have a simple form. The optimal solution would be to compute an index independently for each segment \( x \) morph combination. Call this index the Gittins’ index, \( G_{rmn} \). Outcomes are optimized if, for the \( n^{th} \) consumer (in segment \( r \)), we assign the morph, \( m^* \), which has the largest index. For a proof of optimality, see Gittins (1979).

But in morphing we do not know a consumer’s segment with certainly. We only infer probabilities. Thus, in practice we use the expected value of Gittins’ index, \( E G_{mn} = \sum_r \Pr(r_n = r | \tilde{c}_{nt}) G_{rmn} \) to choose the best morph. Krishnamurthy and Mickova (1999) establish that this intuitive algorithm is close to optimal. Because \( G_{rmn} \to p_{rm} \) as \( n \to \infty \), Gittins’ index converges to the true segment \( x \) morph probabilities. The key difference between the expected-Gittins’-index strategy and the naïve calibration-sample strategy (\( N_{large} \)) is that the expected Gittins’ strategy (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 (about 4,300 annual reviews). 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 and placed on less-valuable web pages. Morphing provides a strategy by which CNET can improve upon context-matching and, hence, provide higher value to its customers. Morphing tells CNET how to communicate while context matching (or behavioral targeting) tells CNET to whom to communicate. 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 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. (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.)

We tested eight potential banners which varied on characteristics likely to appeal differentially to consumers with different cognitive styles. In the control cell, the banners were as-
signed randomly. In the test cell, the morphing algorithm inferred each consumer’s cognitive style and learned the best banner for each cognitive-style segment. Figure 2 provides the banners.

4.2. Cognitive Styles and Banner Characteristics

We first identified a candidate set of scale items from the literature (HULB and references therein; Novak and Hoffman 2009). Using the Greenfield Online panel for a pre-study, we asked 199 consumers to rate themselves on these scales. Factor analysis and scale purification identified eleven items likely to categorize CNET consumers. (Detailed scales and pre-study analyses are available from the authors.)

For a priming study we recorded the clickstreams of 1,292 CNET users and invited them to complete a short priming questionnaire using the eleven purified scale items. We factor analyzed the scale items to identify 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 ipsative cognitive-style dimensions with reliabilities of 0.75, 0.66, and 0.57, respectively. CNET felt they could target most effectively consumer segments that varied on the two most-reliable cognitive-style dimensions. Following HULB’s methods, the morphing algorithm used four cognitive-style segments defined by median splits (2x2: impulsive vs. deliberative x analytic vs. holistic).

CNET’s agency developed the eight banners, seven of which were targeted to cognitive styles. Because morphing automates morph-to-segment matching, the banner designers did not have to make the match themselves; they just had to assure sufficient variation among the banners. (The eighth banner, which did not end up best for any segment, was unconstrained by cognitive styles.) The banners varied on action-links ("learn more" vs. "get it now") and level of detail provided (list of features vs. links to video, etc.). CNET designers also varied the banners on characteristics they judged might appeal to different cognitive styles: font size, font color, back-
ground color, location, and shape. See Figure 2. CNET believed the eight banners had sufficient variation in potential appeal, sufficient for the morphing algorithm to find the best banner(s) for each consumer segment.

[Insert Figure 2 about here.]

4.3. Segment-Specific Click Preferences and Estimation of Segment-Membership

We decompose every click alternative into a vector of 22 click characteristics including dummy variables for areas on the homepage (carousel, 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.). We estimate segment-specific click-characteristic weights, $\omega_r$, from the clickstreams recorded during the priming study of 1,292 CNET users. The preference weights from the priming study enable us to write the likelihood, $\Pr(\tilde{c}_{nt}|r_n = r)$, for each day-to-day consumer’s clicks on the CNET website. Coupled with prior beliefs about segment membership, this likelihood is sufficient to compute the posterior probability, $\Pr(r_n = r|\tilde{c}_{nt})$, that day-to-day consumer $n$ belongs to segment $r$. Details of the estimation of the $\omega_r$ follow HULB. Parameter values are in Appendix 3.

To address the fact that consumers make repeated visits to CNET, we define consumers as active or not active. An active consumer must make at least five clicks on tracked areas of the website. We use cookies so that updating continues through multiple visits when necessary. Five clicks give us sufficient observations to morph ($t_0 = 5$). In the control 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.
4.4. Modifications to Handle Multiple Sessions

For website morphing HULB did not track repeat sessions, but for banner morphing repeat sessions are common (as tracked with cookies). The same banner might be shown in many sessions. CNET (and AT&T) consider the banner a success if the consumer clicks through in at least one session. We adopt that definition of success. To account for interrelated sessions, we use a strategy of temporary updates and potential reversals.

This is best illustrated with a three-session example. Suppose that a consumer sees the same banner in three sessions and clicks through 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 failure, success, failure. Instead, using CNET’s success criterion, the correct posterior should be 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. To achieve these goals we use the following updating strategy. 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. On 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.

The morphing algorithm requires that we set priors for the segment \( x \) morph click-through probabilities. 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 antic-
4.5. 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 other 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 these banners or interactions with AT&T HTC advertising. Sampling appeared random – we detected no systematic differences in the placement of control banners across estimated cognitive-style segments ($\chi^2_{30} = 15.9, p = 0.98$).

Table 1 summarizes the field-test results. Based on prior theory we expect cognitive-style morphing to be more effective when the banner advertisements are targeted to webpages that are relevant to consumers’ purchasing situations. In particular, we expect morphing to improve outcomes relative to that which can be obtained when banners are matched to the context of the webpage. Interactions of quality and source with relevance (involvement) have a long history in advertising research (Chaiken 1980; Petty, Cacioppo and Schumann 1983) and are consistent with current prescriptive theories of targeting (Iyer, Soberman and Villas-Boas 2005; Kenny and Marshall 2000). To the extent that matching banners to cognitive styles facilitates preference learning, observed interactions between targeting and preference learning might be enhanced by morphing (Lambrecht and Tucker 2011). We expect that cognitive-style matching will have less (or no) effect for low-involvement situations, that is, when banners are not matched to context.

Overall, 116,168 consumers saw 451,524 banners. Of these, 32,084 consumers (27.4%) saw 58,899 banners (13.0%) on context-matched webpages where any smart phone was rated, compared, priced, discussed, or pictured. As predicted, morphing achieves significant and sub-
Morphing Banner Advertising

stantial incremental improvements for banners \( (t = 3.0, p = 0.003) \) and for consumers \( (t = 2.2, p = 0.028) \). These measures almost doubled outcomes (83% and 97% lifts, respectively). To put this \textit{in vivo} impact in perspective, context-matching alone in the control cell did not achieve a significant lift for either banners \( (t = 0.3, p = 0.803) \) or consumers \( (t = 1.4, p = 0.167) \).

[Insert Table 1 about here.]

Table 1 also suggests that gains to morphing requires high consumer involvement. There was no significant lift for banners or consumers when the banners were not on context-matched webpages \( (t = 0.5, p = 0.495 \) and \( t = 1.74, p = 0.081 \), respectively). Interactions between morphing and context-matching were significant for banners \( (\chi^2 = 161.8, p < 0.01 \) and for consumers \( (\chi^2 = 8.2, p = 0.017 \). Although there is drop among non-context-matched banners, that drop is marginally significant at best. Although the interaction matches theory, it is worth investigation with future \textit{in vitro} experiments.

5. Automotive Experiment to Test Matching Morphs to Segments

Because managers are interested in metrics other than click-through rates, we supplement the CNET field experiment with an \textit{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 a focused automotive experiment we abstract from the mechanics of banner morphing (Gittins’ learning) to test whether morph-to-segment matching increases brand consideration and purchase likelihood as well as click-through rates. This focused experiment enables us to test the fundamental behavioral hypothesis motivating banner morphing; the hypothesis that banner advertisements are more effective when targeted to consumer segments that vary on cognitive styles. Because consumer involvement interacts with cognitive-style morphing, we also target banners based on buying stage and body-type prefe-
Brand consideration and purchase likelihood are more-intrusive measures than click-through rates. To obtain these measures 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 is not feasible with this design, we use longitudinal methods to effect matching. See Figure 3. In Phase 1, consumers rate all test and control advertisements for their buying-stage and preferred body-type. Two weeks later in Phase 2, consumers answer a series of scales that enable us to assign consumers to cognitive-style segments. In Phase 2 we also obtain pre-measures of brand consideration and purchase likelihood. Phases 1 and 2 replace the morphing algorithm with *in vitro* measurement. These phases enable us to assign each consumer to a segment and to identify the best banners for each segment. The experiment occurs two and one-half weeks later in Phase 3 (4½ weeks after rating banners), when consumers see banners while exploring an automotive information-and-review website. In the test cell banners are matched to cognitive styles (plus buying stage and body-type preference) while in the control cell banners are matched only to body-type preference. (Note that this experiment also extends the definition of consumer segment to include buying stage.)

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,
Morphing Banner Advertising

Cars.com, and AutoTrader, play a major role 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 of banner advertising for their Chevrolet brand. General Motors’ managerial motivation matched our scientific desire to test the premise of morph-to-segment matching.

We created a website that simulated consumer experience on information-and-recommendation websites. Figure 4 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 as on Edmunds.com where body-type preference – coupe, convertible, sedan, SUV, etc. – is displayed prominently on the landing page. Body-type targeting enhances external validity and builds on the interactions identified in the CNET field experiment, but otherwise has no effect on the test-vs.-control comparisons.

[Insert Figure 4 about here.]

5.2. Cognitive Styles and Stage of the Automotive Buying Process

Body-type preference and the automotive buying process stage were measured in Phase
Morphing Banner Advertising

1; cognitive styles were measured in Phase 2. We defined 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 scales adapted from Novak and Hoffman (2009). We factor analyzed their answers to identify three factors. 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. We selected the first two cognitive dimensions to define $2 \times 2$ consumer segments based on mean splits.  

5.3. Test and Control Banner Advertisements

We created test banners that varied on graphics, amount of content, and format to span the cognitive-style segments for each body type and buying stage. The designers sought to provide sufficient variation so that we could target the banners to each cognitive-style segment. Collection-targeted banners emphasize information; comparison-targeted banners compare targeted vehicles to competitors; and commitment-targeted banners stress test drives, finding a dealer, and

---

2 Despite differences in the underlying scales, the type of consumer, and the buying context, the cognitive dimensions for high-tech consumers and automotive consumers are not dissimilar. For each set of consumers, one dimension is impulsive vs. deliberative. The other dimension is either analytic vs. holistic (high tech) or rational vs. intuitive (automotive). More experience might identify common dimensions that can be used across applications.

3 In the automotive experiment we 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.
purchase details. In total there were 75 test banners: (five designs to appeal to different cognitive styles) \( \times \) (three information content designs to appeal to different stages of the buying process) \( \times \) (five body types for targeting).

There were ten control banners: two banner advertisements 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 morph-to-segment-matching experiment. Figure 5 shows the two control banners and all fifteen test banners for one body type (five test banners for every buying stage).

This control was most relevant to General Motors’ business decisions, but if we are to use it as a scientific control we must establish it is 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.

We use the Phase 1 banner evaluations to compare potential test and control banners on meaningfulness, relevance, information content, and believability. Using Phase 1 measures we create an average score that combines the meaningfulness, relevance, information content, and believability ratings. The average score for a test banner is 3.36 (out of 5); the average score for a control banner is 3.70. The combined control banners are significantly larger than a randomization of test banners \( (t = 10.3, p < 0.01) \). Even if we were to use only the two best test banners
for all consumers, the average 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 banners then it is highly likely that morph-to-segment matching will be superior to either a randomly-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 website called “Consumer Research Power.” Consumers search naturally as if they were gathering information for a potential automotive purchase. They do 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 segment where best was defined by the average over consumers in a cognitive-style segment of the Phase 1 measures (meaningfulness, relevance, information content, and believability). Test banners were also targeted by buying stage. Control consumers received banners that alternate between the two control Chevrolet banners.\(^4\) All banners, both test and control, were targeted by body-type preference.

After consumers complete their search on “Consumer Research Power,” we measure Chevrolet brand consideration and purchase likelihood. Dependent measures include click-through rates for banners, click-through rates per consumer, post-measures of brand consideration and purchase likelihood, and the difference in brand consideration and purchase likelihood

\(^4\) 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.
between the post-measures (after Phase 3) and the pre-measures (during Phase 2).

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 is relatively short (five minutes) and occurs 4½ weeks before the Phase 3 experiment. In Phase 1 consumers are not allowed to click through on the banners and, hence, do 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 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 each consumer sees two body-type-targeted banners in the test cell and two body-
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 to be either an equal or sole decision maker in automotive purchases and plan 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 into account.

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 2 reports the post-only comparisons for the morph-to-segment-matching experiment. As in the CNET field experiment (on targeted 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).

[Insert Table 2 about here.]

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

We increase statistical power by accounting for the pre-measures (as in differences of differences) and for variation in segment membership or demographics due to stochastic variation
in random assignment. Table 3 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 3 about here.]

In Table 3 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 2 and 3 suggest that morph-to-segment matching increases brand consideration and purchase likelihood (for automotive consumers) as well as click-through rates. When combined with the CNET field experiment, the focused automotive experiment suggests that the effectiveness of banners is improved when morphing targets banners to consumer segments that vary on cognitive styles and behavioral-context characteristics (buying stage). Both experimental results (CNET and automotive) reinforce the priming-study-based simulations reported in HULB for website morphing.

6. Implications and Future Directions

Online morphing is a nascent technology for improving the effectiveness of banner advertising. The original website morphing paper established the potential for increasing sales if websites morphed their look and feel, but the evaluation was based on data generated in a prim-
Morphing Banner Advertising

ing study. A subsequent field application tested website morphing for a Suruga Bank website (Hauser, Urban and Liberali 2011). The Suruga Bank results were promising, but the small sample was not sufficient to test fully whether the expected Gittins’ index would converge to the best morph.

This paper provides experimental data for morphing banner advertisements. The CNET field experiment establishes that an expected-Gittins’-index strategy enables a website to learn automatically the best morph for each consumer segment. Click-through rates improve substantially for context-matched webpages on a high-traffic website. Theory suggests incremental improvement beyond that obtained by behavioral targeting, but empirical tests remain future research. The automotive experiment tests the fundamental premise of matching morphs to consumer segments. Morph-to-segment matching improves click-through rates, brand consideration, and purchase likelihood. (The automotive experiment assumes the CNET result that, in steady state, the expected-Gittins’-index strategy would identify the best morph-to-segment matches.)

The expected-Gittins’-index provides near optimal learning; we know of no better strategy. By the principle of optimality, the expected-Gittins’-index strategy is superior to a strategy that relies on setting aside the first $N_{large}$ consumers for a random-assignment experiment. Given the high traffic on these websites and the low click-through rates, the improvement 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. In our experiments that cost was the order of $250,000. For high-
volume brands, as in our tests, the incremental improvements in click-through rates, consideration, and purchase intentions justifies the fixed cost. For smaller brands 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 Priming Studies

Today morphing still requires a priming study to (1) establish the definitions of consumer segments and (2) obtain data on click preferences for each segment (the \( \omega_r \)). We envision future applications that rely on norms rather than priming studies. For example, in the applications to date the definitions of the cognitive-style segments are somewhat similar. With more applications, we might use meta-analyses to stabilize cognitive-style definitions so that they might be used without a priming study. Similarly, meta analyses might provide strong priors for segment-based click-characteristic preferences, \( \omega_r \). We might also identify the click-alternative characteristics that best distinguish consumer segments. Such empirical generalizations would enable an advertiser to rely on norms or an abridged priming study. This diffusion of knowledge has already taken place in pre-test markets for consumer packaged goods. Initial studies explored the methods, but later studies built the normative databases. Today, most new product forecasts 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-
Morphing Banner Advertising

lied on experienced professional designers to develop banners or websites to match consumer segments. Morphing implementation identified the best banners for each segment which often spurred further creative development. As we gain more experience we expect that scientific studies will lead to greater insight into the design challenge. Such studies are fertile grounds for new research. The other practical challenge is transportable code. All code has been specific to the application (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.
References


Morphing Banner Advertising


PricewaterhouseCoopers. 2011. IAB Internet advertising revenue report.


### Table 1
**CNET Field Test of Banner Advertisement Morphing**

| Sample Size | Context-matched webpages | | Non-context-matched webpages | |
|-------------|--------------------------|----------------|-----------------------------|----------------|-----------------
|             | Test                      | Control        | Test                        | Control        | Lift          | Significance |
| All banners | 40,993                    | 17,906         | 0.307<sup>b</sup>          | 0.168          | + 83%        | 0.003        |
| Per consumer| 22,376                    | 9,708          | 0.250<sup>b</sup>          | 0.127          | + 97%        | 0.028        |
| All banners | 262,911                   | 129,714        | 0.151                       | 0.160          | - 6%         | 0.495        |
| Per consumer| 59,362                    | 24,722         | 0.144<sup>c</sup>          | 0.197          | - 27%        | 0.081        |

<sup>a</sup> Click-through rates are given as fractions of a percent, e.g., 0.307 of 1%.

<sup>b</sup> Test cell has a significantly larger click-through rate than control cell at the 0.05 level or better.

<sup>c</sup> Test cell has a marginally significantly smaller click-through rate than the control cell at the 0.10 level.

### Table 2
**Automotive Experiment: Banner Advertisement Morphing (Post-only Results)**

(All banners are targeted by body-type preference.)

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

<sup>a</sup> Click-through rates are given as percents. Consideration is a consider-or-not measure reported as a percent. Purchase likelihood is measured with a five-point scale.

<sup>b</sup> Test cell has a significantly larger at the 0.01 level.
### Table 3
Automotive Experiment: Banner Advertisement Morphing
Controlling for Pre-measures, Segment Membership, and Demographics

<table>
<thead>
<tr>
<th></th>
<th>Click-through Coefficient</th>
<th>All Banners Significance</th>
<th>Click-thru per Consumer Brand Consideration Coefficient</th>
<th>Brand Consideration Significance</th>
<th>Purchase Likelihood Coefficient</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-7.400</td>
<td>&lt; 0.01</td>
<td>-3.559</td>
<td>&lt; 0.01</td>
<td>-4.172</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Test vs. control treatment</td>
<td>1.244</td>
<td>a &lt; 0.01</td>
<td>0.562</td>
<td>b 0.062</td>
<td>0.756</td>
<td>a &lt; 0.01</td>
</tr>
<tr>
<td>Pre-measure</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>3.568</td>
<td>a &lt; 0.01</td>
</tr>
<tr>
<td>Buying-process dummies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collect</td>
<td>0.368</td>
<td>0.388</td>
<td>0.553</td>
<td>0.144</td>
<td>0.787</td>
<td>a 0.023</td>
</tr>
<tr>
<td>Compare</td>
<td>0.762</td>
<td>a &lt; 0.01</td>
<td>0.984</td>
<td>&lt; 0.01</td>
<td>0.510</td>
<td>b 0.062</td>
</tr>
<tr>
<td>Commit</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Cognitive-dimension dummies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rational</td>
<td>0.405</td>
<td>0.112</td>
<td>0.184</td>
<td>0.462</td>
<td>0.538</td>
<td>a 0.027</td>
</tr>
<tr>
<td>Intuitive</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Impulsive</td>
<td>0.187</td>
<td>0.485</td>
<td>0.176</td>
<td>0.496</td>
<td>0.271</td>
<td>0.277</td>
</tr>
<tr>
<td>Deliberative</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Female (vs. Male)</td>
<td>-0.463</td>
<td>b 0.070</td>
<td>-0.329</td>
<td>0.184</td>
<td>0.274</td>
<td>0.259</td>
</tr>
<tr>
<td>Age</td>
<td>0.022</td>
<td>0.052</td>
<td>0.020</td>
<td>b 0.056</td>
<td>0.014</td>
<td>0.184</td>
</tr>
<tr>
<td>Income</td>
<td>0.000</td>
<td>0.905</td>
<td>-0.002</td>
<td>0.429</td>
<td>-0.002</td>
<td>0.482</td>
</tr>
<tr>
<td>Log-likelihood ratio</td>
<td>-367.916</td>
<td>a</td>
<td>-223.039</td>
<td>a</td>
<td>-232.057</td>
<td>a</td>
</tr>
</tbody>
</table>

*Significant at the 0.05 level.  *b Significant at the 0.10 level. Sample size 8,991 impressions or 588 consumers. All equations significant at the 0.01 level. Test vs. control is also significant at the 0.01 level with a differences of differences specification.
Morphing Banner Advertising

Figure 1
Conceptual Diagram of Banner Morphing

Tasks

Priming Study

Observe clicks, measure cognitive styles with established scales, and estimate model to predict cognitive styles from observed clicks in day-to-day operation.

Explore

Infer day-to-day visitors’ cognitive styles from their clickstreams.

Optimally learn best banner for a cognitive style by observing click-throughs.

Exploitation

Profit by providing optimal banners for each cognitive-style segment.

Outcome

Click-alternative preferences (and other priors) for each cognitive-style.

Cognitive-style probability.

Update ability to assign banners to cognitive styles.

Optimal click-through rate given current information.

Day-to-day operation
Morphing Banner Advertising

**Figure 2**

Eight Banner Advertisements Targeted to Cognitive Styles (CNET Field Experiment)

1. HTC Aria™
   - Online only! Free shipping!
   - Learn more

2. HTC Aria™
   - See phone details
   - Watch the video

3. HTC Aria™
   - Online only! Free shipping!
   - Android™ 2.1 OS
   - 3.2 inch display
   - 5 MP camera
   - GPS
   - BLUETOOTH®

4. HTC Aria™
   - Online only! Free shipping!
   - Learn more

5. HTC Aria™
   - Online only! Free shipping!

6. HTC Aria™
   - Online only! Free shipping!

7. HTC Aria™
   - Online only! Free shipping!

8. HTC Aria™
   - Online only! Free shipping!
   - Android™ 2.1 OS
   - 3.2 inch HD display
   - 5 megapixel camera
   - GPS and BLUETOOTH®

All ads had the word ‘refurbished’
Figure 3
Automotive Experiment: Longitudinal Design to Effect Morph-to-Segment Matching

Phase 1
Develop potential banners based on pre-studies.
Screen consumers for target market
Consumers indicate body-type preference.
Consumers indicate 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

Factor analyze cognitive-style scales.
Assign consumers to cognitive-style segments and buying stage. Identify the best “morphs” for each cognitive-style segment. All banners are targeted by body type.

Phase 3 (experiment, four and one-half weeks after Phase 1)
Consumers explore “Consumer Research Power” website.
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 4
Simulated Website for Automotive Experiment Matching Morphs to Segments
(Landing page on the left. One of many subsequent pages on the right.)
Figure 5
Example Test and Control Banner Advertisements for the Automotive Experiment
(The left-most banners are controls. The other columns are five banners 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.)
Appendix 1. Mathematical Summary of Morphing Algorithm

A1.1. Notation

Let $n$ index consumers, $r$ index consumer segments, $m$ index morphs, $t$ index clicks, $j$ index click alternatives. Capital letters indicate totals. Let $c_{ntj} = 1$ if $n$ chooses the $j^{th}$ click alternative on the $t^{th}$ 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 $\tilde{c}_{nt}$ be the vector of the $c_{ntj}$ up to an including $T^{th}$ click, let $\tilde{x}_{jtn}$ be the vector of characteristics of the $j^{th}$ click alternative for the $t^{th}$ click for consumer $n$, let $\omega_r$ be the vector of preference weights for the $\tilde{x}_{jtn}$ for the $r^{th}$ segment, let $Pr_0(r_n = r)$ 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 $\omega_r$ from a priming study in which consumers complete scales to identify their segments and we observe the click alternatives they choose. The estimation is based on a logit likelihood with either maximum-likelihood or Bayesian methods. Details are standard, available in HULB, and not repeated here. For online morphing we know the $\tilde{x}_{jtn}$’s for key click alternatives, hence $\tilde{x}_{jtn}' \omega_r$, which is $n$’s observed utility for the $j^{th}$ click alternative for the $t^{th}$ click. Using the logit likelihood (HULB, p. 211), we obtain the probability that observed clicks are chosen given that the consumer is in segment $r$. Bayes Theorem provides $q_{rn}$. 
Morphing Banner Advertising, Appendices

\[
\Pr(\vec{c}_{nT}|\vec{a}_r, \vec{x}_{jkn}s) = \Pr(\vec{c}_{nT}|r_n = r) = \prod_{t=1}^{T} \prod_{j=1}^{J_k} \left( \frac{\exp[\vec{x}_{jtn}^\prime \vec{a}_r]}{\sum_{l=1}^{L} \exp[\vec{x}_{jtn}^\prime \vec{a}_r]} \right)^{c_{ntj}}
\]

(A1)

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

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 \( \alpha_{rmn} \) and \( \beta_{rmn} \). 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 pseudo-likelihood updating:

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

(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. Prior failures did not change the \( \alpha_{rmn} \)'s for each \( r \), so we now add \( q_{rn} \). When a failure becomes
a success, we undo the update that was added to the $\beta_{rmn}$’s for each $r$. Earlier failures caused us to add $q_{rn}$ for each $r$ to the $\beta_{rmn}$’s, hence we now subtract $q_{rn}$ for each $r$ from the $\beta_{rmn}$’s.

**A.1.4. Calculating the Gittins’ Index 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 long-run 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 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 $\alpha_{rmn}$ and $\beta_{rmn}$. ($a$ is fixed.)

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

**A.1.5. Choosing the Morph in Each Period**

When segment assignments are stochastic, we chose the morph in each period that has the highest value for the expected Gittins’ index, $EG_{mn}$. This is the (near) optimal strategy to minimize opportunity loss.

\[
EG_{mn} = \sum_{r=1}^{R} q_{rn} G_{rmn}(\alpha_{rmn}, \beta_{rmn})
\]
Appendix 2. Factor Loadings Matrices for CNET and Automotive Experiments

We factor analyze consumers’ self-evaluations on cognitive-style items 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 priming study consumers are assigned to segments based on median splits (CNET) or mean-splits (automotive) of sum scores.

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

<table>
<thead>
<tr>
<th>Statement</th>
<th>Impulsive vs. Deliberative</th>
<th>Analytic vs. Holistic</th>
<th>Instinctual vs. Not</th>
</tr>
</thead>
<tbody>
<tr>
<td>I rely on my first impressions.</td>
<td>0.086</td>
<td>0.208</td>
<td>0.654</td>
</tr>
<tr>
<td>I am detail oriented and start with the details in order to build a complete picture.</td>
<td>-0.711</td>
<td>-0.066</td>
<td>-0.057</td>
</tr>
<tr>
<td>I find that to adopt a careful, analytic approach to making decisions takes too long.</td>
<td>-0.005</td>
<td>0.699</td>
<td>0.166</td>
</tr>
<tr>
<td>I go by what feels good to me.</td>
<td>-0.055</td>
<td>0.289</td>
<td>0.680</td>
</tr>
<tr>
<td>When making a decision, I take my time and thoroughly consider all relevant factors.</td>
<td>-0.794</td>
<td>-0.098</td>
<td>0.067</td>
</tr>
<tr>
<td>I do not like detailed explanations.</td>
<td>0.220</td>
<td>0.570</td>
<td>0.173</td>
</tr>
<tr>
<td>I reason things out carefully.</td>
<td>-0.748</td>
<td>-0.139</td>
<td>0.000</td>
</tr>
<tr>
<td>Given enough time, I would consider every situation from all angles.</td>
<td>-0.747</td>
<td>-0.034</td>
<td>0.061</td>
</tr>
<tr>
<td>I do not tackle tasks systematically.</td>
<td>0.058</td>
<td>0.753</td>
<td>0.047</td>
</tr>
<tr>
<td>I use my instincts.</td>
<td>-0.100</td>
<td>-0.033</td>
<td>0.798</td>
</tr>
<tr>
<td>I do not approach tasks analytically.</td>
<td>0.108</td>
<td>0.759</td>
<td>0.103</td>
</tr>
</tbody>
</table>
### A2.2. Cognitive-Style Factor Loadings for Automotive Three-Phase Experiment

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

We follow the procedures detailed in HULB to estimate click-characteristics preferences, which were used in the CNET experiment to compute the posterior estimates of cognitive styles in real-time. Table A3.1 shows the maximum likelihood estimates of $\vec{\omega}_r$. This estimation explains 60.5% of the uncertainty ($U^2$ [pseudo-$R^2$] of 0.605).

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

<table>
<thead>
<tr>
<th>Expect the linked page to have pictures or graphs</th>
<th>Constant</th>
<th>Impulsive vs. Deliberative</th>
<th>Analytic vs. Holistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expect the linked page to be focused on a specific question (technical)</td>
<td>-3.947 $^a$</td>
<td>-1.120 $^a$</td>
<td>1.351 $^a$</td>
</tr>
<tr>
<td>Expect the linked page to have large amount of data</td>
<td>1.181 $^a$</td>
<td>0.095</td>
<td>-0.221</td>
</tr>
<tr>
<td>Navigation Bar</td>
<td>6.931 $^a$</td>
<td>-2.459 $^a$</td>
<td>2.349 $^a$</td>
</tr>
<tr>
<td>Carousel</td>
<td>3.946 $^a$</td>
<td>0.190</td>
<td>0.665 $^b$</td>
</tr>
<tr>
<td>More Stories</td>
<td>5.208 $^a$</td>
<td>1.053 $^a$</td>
<td>0.808 $^a$</td>
</tr>
<tr>
<td>Promotion Bar</td>
<td>5.762 $^a$</td>
<td>-1.853</td>
<td>2.630 $^b$</td>
</tr>
<tr>
<td>Popular Topics</td>
<td>3.818 $^a$</td>
<td>1.517 $^a$</td>
<td>-0.981 $^a$</td>
</tr>
<tr>
<td>Tabs</td>
<td>-14.585</td>
<td>1.236</td>
<td>-0.032</td>
</tr>
<tr>
<td>Inside CNET</td>
<td>5.036 $^a$</td>
<td>2.597 $^a$</td>
<td>-0.858 $^b$</td>
</tr>
<tr>
<td>Search category</td>
<td>3.706 $^a$</td>
<td>-2.856 $^a$</td>
<td>2.818 $^a$</td>
</tr>
<tr>
<td>Product-specific reviews</td>
<td>3.741 $^a$</td>
<td>-2.299 $^a$</td>
<td>2.083 $^b$</td>
</tr>
<tr>
<td>Social Influences: expert opinion (&quot;CNET says&quot;)</td>
<td>3.360 $^a$</td>
<td>1.322 $^a$</td>
<td>-1.226 $^a$</td>
</tr>
<tr>
<td>Social Influences: consumer opinion(&quot;what other do&quot;)</td>
<td>2.087 $^a$</td>
<td>0.768 $^a$</td>
<td>-0.237</td>
</tr>
<tr>
<td>Tech-savvy</td>
<td>0.263 $^b$</td>
<td>0.036</td>
<td>-0.176</td>
</tr>
</tbody>
</table>

$^a$ Significant at the 0.05 level.  $^b$ Significant at the 0.10 level.