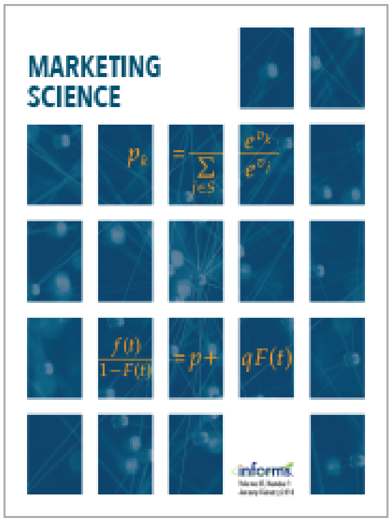

Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Consumer Learning of New Binary Attribute Importance Accounting for Priors, Bias, and Order Effects

Mathew B. Chylinski, John H. Roberts, Bruce G. S. Hardie,

Marketing Science
Vol. 31, No. 4, July–August 2012, pp. 549–566
ISSN 0732-2399 (print) | ISSN 1526-548X (online)
http://dx.doi.org/10.1287/mksc.1120.0719

Consumer Learning of New Binary Attribute Importance Accounting for Priors, Bias, and Order Effects

Mathew B. Chylinski
Australian School of Business, University of New South Wales, Kensington, New South Wales 2052, Australia,
m.chylinski@unsw.edu.au

John H. Roberts
Australian National University, Canberra, ACT 0200, Australia; and
London Business School, London NW1 4SA, United Kingdom,
john.roberts@anu.edu.au

Bruce G. S. Hardie
London Business School, London NW1 4SA, United Kingdom,
bhardie@london.edu

This paper develops and calibrates a simple yet comprehensive set of models for the evolution of binary attribute importance weights, based on a cue–goal association framework. We argue that the utility a consumer ascribes to an attribute comes from its association with the achievement of a goal. We investigate how associations may be represented and then track back the relationship of these associations to the utility function. We explain why we believe this to be an important problem before providing an overview of the extensive literature on learning models. This literature identifies key phenomena and provides a foundation for our modeling of binary attribute importance learning, which can test for three departures from "rational" learning—bias, existence of priors, and the unequal weighting of sample observations (order effects). We apply our models in a laboratory setting under a number of different relationship strengths, and we find that, in our application, consumers' learning about attribute–goal associations exhibits bias and the effects of prior beliefs when the sample realizations occur with and without noise, and order effects when the sample realizations occur with noise. We provide an example of how our models can be extended to learning about more than one attribute.

Key words: consumer utility; preference dynamics; associative learning
History: Received: November 1, 2010; accepted: March 7, 2012; Preyas Desai served as the editor-in-chief and Teck Ho served as associate editor for this article. Published online in *Articles in Advance* May 22, 2012.

1. Introduction

In many markets discontinuous change is common with the introduction of new products, features, and attributes (e.g., Sood and Tellis 2005). The specific case of binary product features is an important one, both in practice and in terms of academic research. Binary attributes may arise from technological standards. NTSC/PAL, HTML5, MP3, and Blu-ray are a few of the many.[1] They may take the form of product variants (e.g., low-calorie beer, diet cola, decaffeinated coffee) or attributes that are naturally dichotomous (electronic fuel injection, cabriolet, rear-screen demister). Binary attributes may occur as ingredient brands or endorsers (e.g., the Heart Foundation Tick, specific country of origin indicators, Dolby Sound, Intel Inside, individual operating systems such as Android). Finally, many manufacturers dichotomize inherently continuous attributes to simplify the consumer decision process. For example, instead of quality, manufacturers may use economy or deluxe; instead of tiers of service level in airlines, they may use coach or business class, and instead of an octane rating for gas, they may use standard or super. When confronted by these binary attributes or features for the first time, consumers have to learn whether their presence will add value to the product.

Turbulent markets with changing products and product features require dynamic models to represent the evolution of sales over time and the behavioral states of consumers that underpin them. Approaches taken in prior research include dynamic coefficient models (Bass et al. 2007), structural models (Gönül and Srinivasan 1996), linear learning models

[1] NTSC: National Television System Committee, PAL: phase alternate line, HTML5: hypertext markup language 5, MP3: MPEG audio layer 3.

(Kuehn 1958, Lilien 1974), and Bayesian models (Iyengar et al. 2007).

Dynamics may be studied at the aggregate or individual level. We may be interested in changes in the *levels* of different variables or, more fundamentally, in the changing *relationship* between them. For example, in terms of promotional response, Guadagni and Little (1983) look at how choice changes with promotional activity, whereas Jedidi et al. (1999) study how the parameters of that relationship change over time.

We are interested in representing individual-level learning processes about the utility of new binary product attributes (also known as importance weights)—that is, the relationship between an attribute and its utility. Understanding the dynamics of importance weight formation is of considerable concern. The behavioral literature suggests that these cannot be taken as given (e.g., Bettman et al. 1998), and in categories with new radical product introductions, managers need to understand the trajectory by which consumer preferences will evolve.

In terms of academic research, the evaluation of binary attributes is a major area of study (e.g., Keeney and Raiffa 1976), whereas behavioral learning experiments in psychology rely mainly on the use of binary features (e.g., van Osselaer and Alba 2003). In studies of consumer decision processes, binary attributes have also attracted considerable attention (e.g., Fader and McAlister 1990).

Given the significance of binary attribute importance evolution, theoretically and in the marketplace, a way of representing the process algebraically is valuable. With it, we can estimate the ultimate preference level of a product with a given set of new attributes and calibrate the time trajectory by which that preference will be formed.

We first review the behavioral learning literature to identify the phenomena involved and examine existing models used to represent the consumer learning process. We then build on previous modeling approaches to develop a simple approach to test for the presence of previously identified phenomena in the learning of associative relationships and hence utility. Finally, we apply our models to consumer learning about a new software agent in a laboratory setting. We examine how the rate, degree, and consistency of learning vary as a function of the association between the attribute's presence and the expected reward it provides, and we test the models' predictive ability. We demonstrate the application of our methodology in the presence of correlated attributes and using a different data generation process before closing with a discussion of the implications and limitations of our research.

We aim to make a contribution at a number of levels. First, and most important, we provide a parsimonious set of models of associative learning that can test for a wide range of behavioral phenomena. Second, we explicitly model the relationship between consumer learning about associations and utility. Finally, we use our approach to empirically examine how learning varies as a function of the strength of the relationship between the presence of an attribute and goal achievement and the presence of noise in one particular application.

2. Consumer Learning of Associations and Relationships

Research in psychology and marketing suggests that consumers have difficulty evaluating stimuli with which they have little or no familiarity. For example, Alba and Hutchinson (2000) show that for unfamiliar attributes, consumers may not have the necessary information to make an accurate estimate of the association between the presence of the attribute and achievement of a goal, whereas Simonson and Tversky (1992) suggest that new features alter the information context facing consumers. Both literatures point to consumers undergoing a learning process whereby information about an attribute is integrated into their belief structures progressively (e.g., Hoeffler and Ariely 1999). van Osselaer and Janiszewski (2001) argue that integration of new information can be represented by an associative learning model, a position supported in psychology (Wasserman and Miller 1997). The associative learning model examines the extent to which the presence of a stimulus or cue (in our case, an attribute) aids the achievement of a consumer's goals (Lawson 1997). Consumers predict the likelihood of goal achievement by integrating information about the co-occurrence of a stimulus and goal achievement into an associative judgment.

Associative learning occurs when a goal, Y, and a stimulus, x, become linked when repeatedly presented together (Wasserman and Miller 1997). We denote goal achievement by $Y = 1$ (and its converse by $Y = 0$) and the presence of a stimulus (in our case, a product attribute) by $x = 1$ (and its absence by $x = 0$). We represent the strength of association between the goal (Y) and stimulus (x) by the conditional probability $\theta_{y|x}$ (i.e., the probability that the goal is achieved ($Y = 1$) or not ($Y = 0$) in the presence ($x = 1$) or absence ($x = 0$) of the stimulus). Lopez et al. (1998b) argue that repeated pairing of goal achievement with a stimulus leads to formation of a perceived association between the two. In the limit, under rational updating, this process allows consumers to estimate the association as the relative frequency with which the goal is achieved in the presence of the attribute.

2.1. Departures from Frequency-Based Estimation

Whereas associative learning provides a conceptual framework for how consumers discover the effect of an attribute, numerous studies indicate that consumer judgments often depart from simple frequency-based estimation in systematic ways. We consider three key departures in the literature relevant to the learning of importances: bias, existence of priors, and unequal weighting of sample observations (order effects).

Biased Learning. Alloy and Tabachnik (1984) claim that learning towards true conditional probabilities may not always occur. In ambiguous environments, perceptual encoding may systematically distort the learning of true associations, leading the consumer to an equilibrium estimate not equal to the true conditional probability. Incoming information may be incorporated in a biased way by confirmatory reasoning, skewed attention/perception effects, anchoring, and other distortions (Hoeffler et al. 2006). These biases may be resistant to disconfirmation.

Prior Beliefs. For a totally new stimulus with no information cues, we would expect the consumer to have highly diffuse priors about the value it will provide in meeting her goals (which is equivalent to assuming that she applies frequency-based estimation). However, the behavioral literature suggests that consumers often enter a learning situation with prior beliefs. They often make use of secondary indicators, such as store experience or relationships with retailers, as a basis for attribute evaluation. If the introduction of new attributes has historically been associated with product improvements (or declines), the availability heuristic may lead to the formation of such priors. Similarly, inferencing might lead consumers to form prior beliefs based on the name or context of an attribute (e.g., Kardes et al. 2004). A range of other evaluation processes or assumptions may also lead the consumer to form prior beliefs about the utility of the new stimulus (e.g., consumer traits such as optimism or pessimism).

Order Effects. Associative learning experiments commonly find order effects. That is, the order in which information is presented to subjects influences their ultimate learning. Recency effects (where later information is given more weight than earlier data) have been observed in a number of associative learning experiments (e.g., Lopez et al. 1998b). Other studies have observed primacy effects (e.g., Marsh and Ahn 2006); that is, once early information has shaped beliefs, later information has difficulty in dislodging these "sticky priors" and is not given equal weight (Hoch and Deighton 1989).

2.2. Models of Consumer Learning

These departures from "rational" decision making in the estimation of associations may appear in concert, with two or three being observed in a given situation. They are likely to be more pronounced in unfamiliar categories (Kahn and Meyer 1991). Thus, our modeling of binary attribute importance formation needs to account for dynamics, including the possibility of biased learning, prior beliefs, and order effects.

Given the importance of associative learning, it is not surprising that there is substantial literature on how to represent it in mathematical terms. Most are at the aggregate level (e.g., McGough 2003), but there are some individual-level models (e.g., Johar et al. 1997). We are specifically interested in consumer learning about new binary attribute importance, but we can build on these models. In particular, to represent importance weight evolution, we can use models of stimulus–goal conditional probability learning, as long as we have a mechanism by which to transform those probabilities to utilities.

Table 1 provides a representative selection of individual-level learning models, with the phenomena each incorporates and the type of learning studied. The research of Rescorla and Wagner (1972) describes an early approach to modeling associative learning in psychology, one in which an error reduction (or partial adjustment) model is used to represent association updating. Pearce (1994) and Cheng (1997) represent two additional foundational papers. Subsequent models have postulated different learning mechanisms and formulations. In particular, a number of authors have represented the learning phenomena identified in the previous section. For example, Anderson and Lebiere (1998) provide a model with prior beliefs, while Gerber and Green (1999) allow for

Table 1 **Phenomena Addressed by Alternative Learning Models**

	Phenomena explained			
	Prior	Bias	Recency	Dependent variable
Rescorla and Wagner (1972)				*Association*
Pearce (1994)				*Recall*
Cheng (1997)				*Probability*
Anderson and Lebiere (1998)	✓		✓	*Recall*
Lopez et al. (1998b)			✓	*Probability*
Cason and Friedman (1999)	✓		✓	*Transaction*
Gerber and Green (1999)	✓	✓		*Probability*
Janiszewski and van Osselaer (2000)				*Association*
Schooler et al. (2001)	✓		✓	*Probability*
Bradlow et al. (2004)	✓		✓	*Utility*
Griffiths and Tenenbaum (2005)	✓			*Probability*
Wallsten et al. (2005)	✓			*Probability and choice*
Kruschke (2006)	✓			*Probability and choice*
Pleskac (2008)	✓			*Probability and choice*

bias in the ultimate learning level. Still others examine the order effects of information (e.g., Bradlow et al. 2004). However, no studies in this tradition focus on the specific problem of attribute weight learning.

Although the models in Table 1 provide useful insight into the dynamics of consumer learning, they suffer a number of disadvantages for the specific problem of describing importance weight formation. First, no one is complete with respect to the phenomena identified in the previous section. Second, many of them are quite complex and difficult to calibrate (often because they were developed for other purposes). Finally, in most cases their algebraic form has been chosen to fit the data. We would ideally like a model or set of models that is based on how the consumer would behave if she used a process of formal statistical inference to estimate importance and to then model departures from that process.

3. A Parsimonious Modeling Approach to Binary Attribute Learning

Our goal is to develop models that are parsimonious with respect to data requirements, fit the learning process well, are complete with respect to the phenomena previously observed in the literature, and are consistent with a utility-maximizing consumer. Our approach to associative learning allows for a wide range of observed phenomena, and it has a ready translation into utility. We know of no other model that captures bias, priors, and order effects simultaneously, nor any that draws the link between associative learning and utility-maximizing behavior.

We tackle the modeling of consumer learning about new binary attribute importances in two steps. First, in §3.1 we describe how the consumer learns about the relationship between the presence (or absence) of a feature and the achievement of her goals. Second, in §3.2 we show how this cue–goal association may be translated into utility judgments about the value of the attribute to the consumer.

3.1. Learning About the Relationship Between a Feature's Presence and Goal Achievement

To represent how consumers learn about associations, we initially assume a separable utility function so we can study one attribute at a time.[2] The consumer does not know the true association between the presence or absence of an attribute and the achievement of her goal, but she observes the relative co-occurrences over t exposures. That is, the consumer is exposed to information pairs (i.e., sample realizations) $(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_t, y_t)$, where at the ith observation, the attribute is either present ($x_i = 1$) or not ($x_i = 0$) and the goal is achieved ($y_i = 1$) or not ($y_i = 0$).

We are interested in how the consumer's estimates of the associative probabilities $\theta_{y\,|\,x}$ evolve over time. Let $\hat{\theta}^t_{1\,|\,1}$ denote the consumer's estimate at time t of $\theta_{1\,|\,1}$, the underlying probability that the goal is met when $x = 1$.[3] Similarly, let $\hat{\theta}^t_{1\,|\,0}$ denote the consumer's estimate at time t of $\theta_{1\,|\,0}$, the probability that the goal is met when $x = 0$.

Model 1: Base Model Learning. Faced with t information pairs, we assume that a consumer with no information processing bias, no prior beliefs, and no order effects will approach the problem as a statistician. Given the assumed Bernoulli nature of goal achievement in the presence or absence of the focal attribute, the likelihood function for $\theta_{1\,|\,1}$ and $\theta_{1\,|\,0}$ is given by

$$L\big(\theta_{1\,|\,1}, \theta_{1\,|\,0} \mid (x_1, y_1), \ldots, (x_t, y_t)\big) = \prod_{i=1}^{t} \theta_{1\,|\,1}^{x_i y_i} (1-\theta_{1\,|\,1})^{x_i(1-y_i)} \theta_{1\,|\,0}^{(1-x_i)y_i} (1-\theta_{1\,|\,0})^{(1-x_i)(1-y_i)}.$$

It follows that her maximum likelihood estimate of $\theta_{1\,|\,1}$ is given by

$$\hat{\theta}^t_{1\,|\,1} = \frac{\sum_{i=1}^{t} x_i y_i}{n^t_{\bullet\,|\,1}}, \tag{1}$$

where $n^t_{\bullet\,|\,1} = \sum_{i=1}^{t} x_i$ is the number of trials out of t for which the attribute is present. (The maximum likelihood estimate of $\theta_{1\,|\,0}$ is $\hat{\theta}^t_{1\,|\,0} = (\sum_{i=1}^{t}(1-x_i)y_i)/n^t_{\bullet\,|\,0}$, where $n^t_{\bullet\,|\,0} = \sum_{i=1}^{t}(1-x_i)$ is the number of trials out of t for which the attribute is missing. We omit estimates of $\theta_{1\,|\,0}$ going forward because they follow directly by analogy to $\theta_{1\,|\,1}$.) For notational simplicity, we denote $\sum_{i=1}^{t} x_i y_i / n^t_{\bullet\,|\,1}$, the observed frequency of goal achievement relative to attribute occurrences, by $f^t_{1\,|\,1}$, and $f^t_{1\,|\,0}$ by analogy.

Model 2: Biased Learning. Model 1 is based on a consumer with no information processing bias. Suppose the consumer systematically processes information with bias because of selective attention, distortion, and/or retention of information, as suggested in the previous section. In particular, we assume that the presence of the attribute ($x = 1$) is correctly perceived. However, errors occur in terms of whether or

[2] See Keeney and Raiffa (1976) for a discussion of this assumption in multiattribute utility models in decision analysis and Blackorby et al. (1977) for an economics perspective. We discuss and demonstrate the extension to more than one attribute, including inter-attribute correlation, in §6.2.

[3] Strictly speaking, "time t" is a sample realization number rather than a reference to chronological time.

not achievement of the goal is correctly observed. The consumer perceives goal achievement $\hat{y}_t$ times in the $n^t_{\bullet|1}$ realizations of $x = 1$. Let $p_t = \hat{y}_t/(\sum_{i=1}^t x_i y_i)$.[4] If the consumer is unbiased, she will perceive $\sum_{i=1}^t x_i y_i$ goal realizations and $p = 1$. Note that $p < 1$ corresponds to the consumer underestimating the association between the attribute and goal realization, and $p > 1$ corresponds to an overestimation. This is consistent with Gerber and Green (1999), who assume that an individual assimilates a fixed proportion of disconfirming information. An alternative, but algebraically similar process, would occur if the consumer does record the observation but fails to update her estimate of the cue–goal association (Miller and Matzel 1988, van Osselaer and Alba 2000).[5]

In expectation, the maximum likelihood estimate of $\theta_{1|1}$ after t trials is given by

$$\hat{\theta}^t_{1|1} = \frac{\sum_{i=1}^t p x_i y_i}{n^t_{\bullet|1}}. \tag{2}$$

$\hat{\theta}_{1|0}$ follows analogously in terms of a bias parameter defined similarly to p.

Model 3: Prior Beliefs Influencing Learning. The above two models of association estimation assume that the consumer has no prior beliefs before gaining information about the efficacy of the attribute in achieving her goal. The literature section suggested that this might not be the case.

Therefore, we allow the consumer to approach the new attribute with priors about the association. Given the Bernoulli process that lies behind Equation (1), we assume that prior beliefs about the probability $\theta_{1|1}$ are captured by a beta distribution with parameters α and β. (Similarly defined prior beliefs about the probability of goal achievement in its absence, $\theta_{1|0}$, are distributed beta with parameters analogous to α and β.)

After t trials the consumer's beliefs about $\theta_{1|1}$ are captured by a beta distribution with parameters $\alpha + \sum_{i=1}^t x_i y_i$ and $\beta + \sum_{i=1}^t x_i(1-y_i)$ (Massy et al. 1970). Letting $\mu = \alpha/(\alpha+\beta)$ and $\delta = \alpha + \beta$ (with $x = 0$ analogs similarly defined), it follows that

$$\hat{\theta}^t_{1|1} = \left(\frac{\delta}{\delta + n^t_{\bullet|1}}\right)\mu + \left(\frac{n^t_{\bullet|1}}{\delta + n^t_{\bullet|1}}\right) f^t_{1|1}, \tag{3}$$

where μ is the mean of the prior distribution with which the consumer enters the learning process, and δ is the effective sample size in Bayesian updating (Roberts and Urban 1988). We note that this learning model with prior beliefs reduces to Model 1 as $\delta \to 0$, corresponding to a consumer with totally diffuse priors.

Insights into the dynamics of learning are obtained via the following rearrangement of Equation (3):

$$\hat{\theta}^t_{1|1} = f^t_{1|1} + \left(\frac{\delta}{\delta + n^t_{\bullet|1}}\right)(\mu - f^t_{1|1}). \tag{4}$$

This suggests that the consumer will deviate from the maximum likelihood estimate by an amount that is the product of the distance that the mean of her prior is from it $(\mu - f^t_{1|1})$ and the proportion of information contained in her prior beliefs $(\delta/(\delta + n^t_{\bullet|1}))$.

We note that if a new attribute is introduced into a product category with which the consumer is familiar, we would not expect to see learning about goal achievement in the absence of the attribute $(\theta_{1|0})$. If the product category is also new, changes in $\theta_{1|0}$ will capture learning about other product elements.

Model 4: Order Effects Model (Learning with Recency or Primacy). To represent the possibility that earlier information is partially forgotten or discounted (recency) or is more influential in cementing beliefs early (primacy), we allow the weight of an observation to be a function of when it is received, replacing x_i by $\lambda_i x_i$ for $x_i = 1$ (and analogously for $x_i = 0$). Following Bradlow et al. (2004), we assume geometric decay (or growth) in the observation's weight. That is, $\lambda_i = \lambda^i$, where λ is the decay/growth rate for $x_i = 1$. Conditional on λ, we can write the maximum likelihood estimator of the order effects model as

$$\hat{\theta}^t_{1|1} = \frac{\sum_{i=1}^t \lambda^i x_i y_i}{\sum_{i=1}^t \lambda^i x_i}. \tag{5}$$

For $\lambda > 1$, recent observations of $x_i = 1$ are more influential (recency), while for $\lambda < 1$, their effect is attenuated (primacy). ($\hat{\theta}^t_{1|0}$ is similarly defined for $x_i = 0$.)

Combined Models. These phenomena may exist in concert. We may combine the phenomena in Equations (2), (3), and (5) (and their $\theta_{1|0}$ analogs) in a reasonably straightforward way to include any combination of bias, priors, and order effects. For example, the complete bias + priors + order effects specification (Models 2 + 3 + 4) states that the consumer's belief about $\theta_{1|1}$ after t trials is given by

$$\hat{\theta}^t_{1|1} = \left(\frac{\delta}{\delta + \sum_{i=1}^t \lambda^i x_i}\right)\mu + \left(\frac{\sum_{i=1}^t \lambda^i x_i}{\delta + \sum_{i=1}^t \lambda^i x_i}\right)\frac{\sum_{i=1}^t \lambda^i p x_i y_i}{\sum_{i=1}^t \lambda^i x_i}. \tag{6}$$

[4] The bias that the subject introduces in estimating the true association between the goal and the attribute is the ratio of her long-term estimate of the association and the true association, $\hat{\theta}^{\infty}_{1|1}/\theta_{1|1}$. As such, it is not time dependent. We have subscripted it by t in this equation because the subject sees $f^t_{1|1}$ rather than $\theta_{1|1}$ (as sample realizations of goal achievement are drawn from a distribution with $\theta_{1|1}$ as their mean). As t increases, any difference between $f^t_{1|1}$ and $\theta_{1|1}$ gets small. For notational simplicity, we drop the time subscript on p in the model development that follows. Note that p is constrained by $p_t \le n^t_{\bullet|1}/\sum_{i=1}^t x_i y_i$ for $\hat{\theta}^t_{1|1}$ to belong to $[0, 1]$.

[5] We are grateful to an anonymous reviewer for this observation.

3.2. Translating an Understanding of Association into Utility Judgments

Our research objective is to present a parsimonious representation of consumers' learning about the importance of a binary attribute. The models presented in Equations (1), (2), (3), (5), and (6) provide representations of how consumers' subjective estimates of the *association* between the presence or absence of an attribute and the achievement of a goal evolve under different assumptions about consumer learning. The next step is to explore the mapping of these associations to the expected utility of goal achievement.

We are interested in the utility of a new binary attribute x in a multiattribute utility setting. Given our assumption of separability, we can consider the contribution of the attribute to the product's overall utility independently of other attributes. We denote the utility of x as $U(x)$.

The standard approach to multiattribute utility theory typically represents $U(x)$ as a linear function of the attribute level x weighted by the consumer's attribute importance or tastes w_1 (e.g., Keeney and Raiffa 1976):

$$U(x) = w_0 + w_1 x. \tag{7}$$

The importance of a binary attribute is simply the change in utility associated with its presence relative to its absence (i.e., $U(x=1) - U(x=0)$).

It is typically assumed that w_1 is an intrinsic property of a consumer, as if part of the consumer's genetic makeup (e.g., McFadden 2001), and that attribute levels are determined independent of the consumer through the value chain of production. Hence, the consumer knows her tastes but may be uncertain about the product's attribute levels. This leads to the modeling of utility dynamics based on consumer learning about attribute levels (e.g., Roberts and Urban 1988).

An alternative approach to multiattribute utility theory draws on the ideas of Lancaster (1966): utility is not derived directly from the attribute itself but from the extent to which it assists in undertaking a desirable consumption activity, which more generally results in the achievement of a goal Y. Lancaster expresses this idea in the following algebraic terms:

$$U(x) = a(x)U_Y,$$

where U_Y is the utility associated with goal achievement (or successful undertaking of the activity), and $a(x)$ is a scalar that represents the extent to which attribute x assists in undertaking that consumption activity. It follows that the utility of a binary attribute is a function of the extent to which its presence assists in undertaking the consumption activity, relative to its absence:

$$U(x=1) - U(x=0) = [a(x=1) - a(x=0)]U_Y. \tag{8}$$

With reference to Equation (7), this gives us

$$w_1 = \Delta a U_Y, \tag{9}$$

where $\Delta a = a(x=1) - a(x=0)$.

This Lancasterian-motivated interpretation of a consumer's attribute importance judgment has implications for modeling a consumer's learning about w_1. Equation (9) decomposes attribute importance into two components, one determined by the consumer's utility of a consumption activity (or goal achievement), U_Y, and the other determined by the degree to which the product attribute facilitates the achievement of the consumption-related goal. We assume that a consumer's uncertainty about the value of the attribute, w_1, arises from her being uncertain about Δa, the degree to which the attribute facilitates the achievement of goal Y. There is no uncertainty about the utility derived from Y. (A natural extension would be to allow for learning about the utility of the consumption activity, U_Y.) Dynamics in attribute importances are represented by adding a time subscript to Equation (9):

$$w_{1t} = \Delta a_t U_Y. \tag{10}$$

Note that under this formulation, changes in utility (and, from them, choice) arise from changes in consumers' attribute importance judgments (w_{1t}), not changes in perceived attribute or feature levels (which are assumed known).

So far, our Lancasterian-motivated view of importance weights assumes a deterministic relationship between the presence or absence of an attribute and the extent to which it assists in a consumption activity. We now introduce a stochastic element and assume that an attribute's presence may change the *probability* (rather than degree) of goal achievement. We propose that the presence of the new attribute leads to a utility of U_Y with probability $\theta_{1|1}$, while its absence leads to a utility of U_Y with probability $\theta_{1|0}$. Denoting expected utility by $EU(\cdot)$, the expected utility associated with the attribute's presence and absence are $EU(x=1) = \theta_{1|1}U_Y$ and $EU(x=0) = \theta_{1|0}U_Y$, respectively. It follows that the utility of a binary attribute is a function of the extent to which its presence or absence changes the probability of goal achievement. Thus, the stochastic analog of the deterministic Equation (9) is

$$w_1 = \Delta\theta_{1|x} U_Y, \tag{11}$$

where $\Delta\theta_{1|x} = \theta_{1|1} - \theta_{1|0}$. Within the psychology and consumer behavior literatures, the difference in associations (i.e., $\Delta\theta_{1|x}$) is known as the *contingency* (Janiszewski and Warlop 1993, Lopez et al. 1998a). It is the increased (or decreased) probability of goal achievement brought about by the presence of the attribute relative to its absence. We can therefore write our expected utility equivalent of Equation (7) as

$$EU(x) = \theta_{1|0}U_Y + \Delta\theta_{1|x}U_Y x. \tag{12}$$

Thus, Equation (12) shows that the consumer's increased valuation on seeing the attribute present is proportional to the differential probability that its presence has on the consumer achieving her goal, $\Delta\theta_{1|x} = \theta_{1|1} - \theta_{1|0}$—the so-called goal contingency of the attribute. Equation (12) establishes the role of associations within an expected utility framework. Combining Equation (12) with the association formation Equations (1)–(3), (5), and (6), we see how rational updating, as well as updating with bias, priors, and order effects, impacts importance weight learning.

We can extend this logic to the case of K attributes $(x^1, x^2, \ldots, x^K)$. Let $\theta_{1|x^1x^2\cdots x^K}$ denote the probability of goal achievement given attribute levels $x^1, x^2, \ldots, x^K$. Assuming the impact of a change in the level of an attribute is independent of the levels of the other attributes (i.e., separability), the multivariate extension of Equation (12) is straightforward:

$$EU(x^1, x^2, \ldots, x^K) = \theta_{1|00\cdots 0}U_Y + \sum_{k=1}^{K} \Delta\theta_{1|x^k}U_Y x^k, \tag{13}$$

where $EU(x^1, x^2, \ldots, x^K)$ is the expected utility given the levels of attributes $x^1, x^2, \ldots, x^K$, $\theta_{1|00\cdots 0}$ is the probability that the goal is achieved when $x^1 = x^2 = \cdots = x^K = 0$, and $\Delta\theta_{1|x^k}$ is the change in $\theta_{1|x^1x^2\cdots x^k\cdots x^K}$ as x^k goes from 0 to 1, holding all other x^j $(j \neq k)$ constant. (We examine this in greater depth in §6.2 for the case of $K = 2$, including the case of a nonseparable utility function.)

4. Model Calibration

We have presented a modeling approach for understanding the evolution of consumer learning of the importance of a new binary attribute. We propose that the utility a consumer ascribes to an attribute comes from its association with the achievement of a goal motivating the consumption of the associated product or service. We have identified three departures from "rationality"—bias, existence of priors, and the unequal weighting of sample observations (order effects)—and presented a set of parsimonious models that capture the evolution of preferences given these departures. We now test these alternative representations and the association-to-utility mapping using data from an experiment based on a classical conditioning design. We examine how learning varies as a function of strength of the relationship between the presence of an attribute and goal achievement, and the presence of noise in one particular application.

Our primary experiment collects data on subjects' utility judgments and assessments of the probability of goal achievement Y. With repeated measurements for each subject, we can estimate how attribute–goal association and attribute–utility judgments are updated with more information.

The masking task is one of stock picking. The subjects' task was to choose the equity from a set of four that they thought would increase in value the most. Thus, the goal (Y) is one of picking the stock that will return the highest gain. Subjects were nominally paid \$100 when they selected an equity with the highest short-term capital gain. There were no gains or losses associated with selecting an equity that did not yield highest short-term capital gain. Therefore, $U_Y = \$100$. This stock-picking task was repeated multiple times, with the set of four stocks varying randomly on each stock selection occasion.

For each stock selection, subjects saw the stock recommendation from a (hypothetical) Internet-based equity trading software agent called E-T.com. This agent had (or did not have) two potentially salient product attributes called "Neural Net" (x) and "Temporally Continuous" (z).[6] The software agent advised subjects on which equity to choose, but they did not have to accept its recommendation.

After each subject had selected the equity that she thought would increase in value the most, the subject was then told which equity actually had, thus providing information about goal achievement and, indirectly, the accuracy of an agent with specific levels of the two attributes x and z. For each selection, the agent with a given set of characteristics recommended the correct stock according to a Bernoulli process. An agent with $x = 1$ recommended the correct stock with probability $\theta_{1|1}$, while an agent with $x = 0$ recommended the correct stock with probability $\theta_{1|0}$. (z was included as a control attribute, and initially, only the presence or absence of x affected the accuracy of the agent's recommendations in terms of subjects' goal achievement.) This provided the subject with information pair (x_i, y_i). The task of equity selection was a mask for the true (conditioning) purpose of the experiment.

For each replication of the experiment, the subject undertook four batches of four stock picks. The four stock pick tasks in each batch were accompanied by recommendations from an agent with given levels of (x, z) pairing. Subjects saw the four batches, associated with the four $(0, 1)$ combinations of the two attributes, in a random order. Having made 16 stock picks, subjects were then asked to estimate how many

[6] While we denote attribute levels by $(x^1, x^2, \ldots, x^K)$ in the K-attribute case, we use x and z in the two-attribute case for notational simplicity.

times the goal would be achieved over the next 100 trades if they followed the agent's recommendation (for an agent with given levels of x and z). We denote these quantities by $N_{jr}(1,1)$, $N_{jr}(1,0)$, $N_{jr}(0,1)$, and $N_{jr}(0,0)$, where, for example, $N_{jr}(1,0)$ denotes subject j's estimate for an agent with characteristics $x=1$ and $z=0$ at the end of replication r.

To get a direct assessment of the utility of an agent with characteristics (x,z), subjects were also asked at the end of each replication how much they would pay (or have to be paid) per trade to use an agent with those characteristics relative to a "baseline" agent (e.g., what would make an agent with characteristics $(x=1, z=1)$ of equal value to an agent with characteristics $(x=0, z=1)$), thus giving an assessment of the utility of attribute x. Each subject undertook 12 replications $(r=1,\ldots,12)$, thereby providing data that enables us to study the dynamics of learning. Examples of the stimuli used and exact measures taken are included in Web Appendix A (at http://dx.doi.org/10.1287/mksc.1120.0719).

To investigate whether the phenomena represented in our models vary as a function of the strength of relationship between the attribute and the goal, we manipulate the contingencies (i.e., $\theta_{1|1}-\theta_{1|0}$) between subjects. We chose nine different contingencies, which are given in Table 2, in terms of their constituent associations. Because the achievement (or not) of the goal was generated using a random draw from a Bernoulli process, the realized conditional probabilities are not exactly the same as the underlying (programmed) values of the goal–attribute associations.

We recruited 135 subjects for the experiment from a population of undergraduate business and MBA students. This allowed 15 subjects per contingency condition. Subjects were offered a sure reward of \$20 plus a chance of winning \$500 to participate. Subjects were randomly assigned to the experimental conditions.

The new product, a software agent to assist in equity selection, is a realistic one and is salient to business students. Many such products exist in the market (see, for example, Stockpickpro (http://stockpickpro.com/) or Stock Assault 2.0 (http://www.stockassault.com/)). Such electronic agents have also attracted considerable research attention (see Xiao and Benbaset 2007 for a review).[7]

Table 2 Conditional Probabilities and Realized Associations in Experiment 1 (Noise Design)

Condition	Programmed conditional probabilities $\theta_{1\|1}$	Programmed conditional probabilities $\theta_{1\|0}$	Realized conditional probabilities $\theta_{1\|1}$	Realized conditional probabilities $\theta_{1\|0}$
1	0.25	1.00	0.25	1.00
2	0.25	0.75	0.26	0.79
3	0.25	0.50	0.26	0.52
4	0.25	0.25	0.26	0.25
5	0.50	0.25	0.51	0.22
6	0.75	0.25	0.82	0.30
7	0.75	0.13	0.82	0.14
8	1.00	0.25	1.00	0.24
9	1.00	0.13	1.00	0.14

4.1. Estimating the Association Models

As previously noted, subjects were asked at the end of each replication for their estimate of how many times their goal would be achieved over the next 100 trades if they followed the recommendations of agents with (x,z) characteristics $(1,1)$, $(1,0)$, $(0,1)$, $(0,0)$. Equations (1), (2), (3), (5), and (6), together with their analogs for $\theta_{1|0}$, specify our representations of individual j's $(j=1,\ldots,J)$ judgment of the probabilities of goal achievement after t exposures for each of our models. We wish to see the degree to which each model describes the evolution of the $N_{jr}(\cdot,\cdot)$.

The models of associative learning provide estimates of the consumer's probabilities of goal achievement given the presence or absence of the attribute after exposure to t information pairs, (x_1,y_1), $(x_2,y_2),\ldots,(x_t,y_t)$. Subjects' direct assessments of these probabilities were solicited at the end of each replication. Since z is an irrelevant attribute for the purposes of this study, we tested whether its presence or absence affected the subject's estimated associations. We found no statistical difference between the two levels of the irrelevant attribute (z), so we pooled the $z=1$ and $z=0$ data to increase the statistical power of our results for the focal attribute x. This means that by replication r, the subject had been exposed to $t=8r$ pairs of information for both $x=1$ and $x=0$ ($4r$ associated with both $z=1$ and $z=0$).

The parameter estimates for each model were obtained by using nonlinear least squares. Given the repeated measures nature of the data collected, we estimate a mixed nonlinear model allowing for a subject-specific random effect (with variance σ_v^2) and decompose the error variance of the model into replication-specific and contingency-specific components (σ_r^2 and σ_c^2, respectively). We accommodate replication-specific error variances by allowing them to vary over replication (r) in the following manner: $\sigma_r^2=\exp(r_0+r_1 r)$. Similarly, we accommodate contingency-specific variances by allowing them to vary as function of the observed relative frequency to date for each contingency. For the case of $x=1$, we

[7] Note that we are not suggesting that the majority of such stock-picking recommendation agents outperform random picks, or even that any do. All that we need is for subjects to believe that it is possible that they may. The presence of such products on the market, together with feedback from subject debriefing, suggests that this is not an unreasonable assumption.

have $\sigma_c^2 = \exp(c_0 + c_1 f^t_{1|1})$, and we use an analogous formulation for the case of $x = 0$ using $f^t_{1|0}$.

We manipulated the degree of association between the attribute presence/absence and goal achievement (see Table 2) so that we can investigate whether the phenomena represented in the models vary as a function of the strength of relationship between the attribute and the goal. In particular, we are interested in whether consumers underlearn ($p < 1$) or overlearn ($p > 1$) when the attribute more frequently leads to goal achievement ($\theta_{y|x}$ is higher), whether the updating of priors occurs more quickly (δ is lower), and whether order effects are more pronounced (λ is higher or lower). Therefore, rather than estimating a common p, δ, and λ, we allow them to vary across conditions, writing them as a function of the observed relative frequency to date for each association. For the case of $x = 1$, we have

$$\left.\begin{array}{l} p_t = p_1 + p_2 f^t_{1|1} \\ \delta_t = \delta_1 + \delta_2 f^t_{1|1} \\ \lambda_t = \lambda_1 + \lambda_2 f^t_{1|1} \end{array}\right\}. \tag{14}$$

(These are similarly defined for $x = 0$ using $f^t_{1|0}$.)

We believe that there is no reason why the prior means should change with the (ex post) observed frequency, and so we do not let μ vary across associations. Note that p, δ, and λ are subscripted by t simply because the observed relative frequencies (drawn from the Bernoulli distribution) may vary from replication to the next.

4.2. Estimating the Utility Model

Our framework suggests that consumers learn about the association between goal achievement and the presence of an unfamiliar attribute, and thus they update their estimate of its importance weight (and thus utility). Under Equation (12) we expect their utility estimates associated with the presence of an attribute (relative to its absence), $EU(x = 1) - EU(x = 0)$, to equal the estimated contingency, $\hat{\theta}_{1|1} - \hat{\theta}_{1|0}$, times the utility of goal achievement, U_Y, which is \$100 in our experiments. To examine the proposed contingency to utility transformation, we fit

$$\widehat{\Delta EU}^t = \tau + \rho(\hat{\theta}^t_{1|1} - \hat{\theta}^t_{1|0}), \tag{15}$$

where $\widehat{\Delta EU}^t$ is the subject's estimate of the change in utility at time t (i.e., their willingness to pay per trade to use an agent with $x = 1$ relative to a baseline agent with $x = 0$ at time t), and $\hat{\theta}^t_{1|1} - \hat{\theta}^t_{1|0}$ is their estimate of the contingency at time t. (We note that the utility model proposed in Equation (12) predicts that $\tau = 0$ and $\rho = U_Y = 100$.) Given the repeated measures nature of the data collected, we estimate Equation (15) using a mixed linear model that also allows for a subject-specific random effect (with variance σ_ν^2) and replication-specific error variances σ_r^2. As with the association regressions, we accommodate replication-specific error variances by allowing them to vary over replication in the following manner: $\sigma_r^2 = \exp(r_0 + r_1 r)$.

5. Results

There are four stages to our main analysis. First, we fit our four models (and the "combined" model represented by Equation (6)) to examine whether we observe departures from rational updating in the noise condition. We then compare the fit of our models to alternative models also motivated by the literature, aiming for at least comparable fit, while maintaining simplicity, completeness, and nesting. Next, we examine the implications of learning about associations and contingencies for importance weight formation, using the utility model in Equation (12). Finally, we test the predictive ability of our models.

5.1. Fitting the Basic Models

The estimation results from fitting Models 2–4 and the combined model are presented in Table 3. We first note that the results associated with Model 1 are not included in the table. Recall that this model assumes that the subject approaches the problem as a statistician and that her perceived associations are therefore the relative frequencies of goal–attribute co-occurrences observed in the data; there are no other model parameters to estimate. Comparing these observed relative frequencies to directly elicited estimates gives a Model 1 log likelihood (LL) of $-14{,}028$ for the case of $x = 1$ ($\theta_{1|1}$) and $-14{,}076$ for the case of $x = 0$ ($\theta_{1|0}$).

For all models in Table 3, we see evidence of random effects for subjects ($\sigma_\nu^2 > 0$). In all cases, variance also decreases as the association increases ($c_1 < 0$) and as the consumer gathers more information over time ($r_1 < 0$). All models provide an improvement in fit over the model of the consumer as statistician. To gain further insight, we examine the models individually.

Biased Learning. Looking at the estimates for Model 2, we see that allowing subjects to learn to a probability of goal achievement other than the true frequency can improve fit. For $x = 1$ for low associations, consumers actually overlearn the value of the attribute ($p_1 > 1$), but this effect decreases with increasing probability ($p_2 < 1$) until finally consumers underlearn ($p = 1$ at $f_{1|1} = (1 - 1.21)/(-0.28) = 0.75$). This seems to be consistent with Kahneman and Tversky's (1979) observation in prospect theory that consumers overestimate low-probability events but underestimate high-probability ones. However, the effect is not observed in the absence of the attribute: for $x = 0$, no significant bias in learning is observed.

Table 3 Estimates of $\theta_{1|0}$ and $\theta_{1|1}$ for Proposed Models of Association (Noise Data)

	Model 2 (Bias)		Model 3 (Priors)		Model 4 (Order effects)		Combined (Bias + Priors + Order effects)	
Parameter	$\theta_{1\|1}$	$\theta_{1\|0}$	$\theta_{1\|1}$	$\theta_{1\|0}$	$\theta_{1\|1}$	$\theta_{1\|0}$	$\theta_{1\|1}$	$\theta_{1\|0}$
σ_v^2	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
c_0	5.53***	5.32***	5.53***	5.27***	5.52***	5.32***	5.51***	5.26***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
c_1	−0.90***	−0.46***	−0.97***	−0.38**	−0.88***	−0.53***	−0.95***	−0.45***
	(0.10)	(0.14)	(0.11)	(0.14)	(0.10)	(0.14)	(0.11)	(0.14)
r_0	6.82***	6.79***	6.45***	6.46***	6.84***	6.82***	6.48***	6.51***
	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)
r_1	−0.67***	−0.67***	−0.53***	−0.58***	−0.67***	−0.66***	−0.55***	−0.61***
	(0.09)	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
p_1	1.21***	1.01					1.17***	0.90*
	(0.06)	(0.05)					(0.06)	(0.05)
p_2	−0.28***	−0.03					−0.21***	0.07
	(0.06)	(0.06)					(0.06)	(0.06)
μ			0.55***	0.52***			0.51***	0.48***
			(0.05)	(0.05)			(0.04)	(0.05)
δ_1			3.50*	7.26***			7.79*	11.70**
			(1.96)	(2.02)			(4.29)	(5.64)
δ_2			1.71	−4.33*			−1.29	0.38
			(2.80)	(2.57)			(5.29)	(10.87)
λ_1					1.31**	1.28***	1.24**	0.99
					(0.13)	(0.09)	(0.10)	(0.14)
λ_2					−0.30	0.11	0.08	1.31***
					(0.25)	(0.23)	(0.18)	(0.46)
LL	−13,387	−13,536	−13,357	−13,487	−13,395	−13,517	−13,334	−13,444
BIC	26,808	27,107	26,752	27,013	26,825	27,069	26,727	26,946

Notes. Standard errors are in parentheses. The significance of p_1 and λ_1 are tested relative to 1.0. The Model 1 log likelihood is −14,028 for $\theta_{1|1}$ and −14,076 for $\theta_{1|0}$; the associated Model 1 BIC is 28,063 for $\theta_{1|1}$ and 28,160 for $\theta_{1|0}$. The number of observations for all models is 3,240.

*Denotes significance at the 10% level; **denotes significance at the 5% level; and ***denotes significance at the 1% level.

Prior Beliefs. Models incorporating prior beliefs also improve fit relative to the base model. Looking at the values of μ in Table 3 (corresponding to Equation (3)), we see that in both cases they are close to 0.5. That is, the subject starts the experiment with approximately a 50/50 prior that the product (with or without the attribute) will allow her to achieve her goal. This finding is remarkably robust, not just across all models in the noise experiment, but with the two additional experiments reported in §6 as well (no-noise and multiple attributes). Also of interest is δ, the strength of priors (in terms of the number of information pairs or equivalent sample size). At $f_{1|1} = 0$, the value of δ_1 suggests that it takes approximately 3.5 pieces of information for $x = 1$ and 7.3 for $x = 0$ for learning to outweigh prior beliefs. (The differences between these two numbers becomes insignificant when the effect of relative frequency δ_2 is considered.) Unlike for priors, the effect of the level of association on learning is not strong, with δ_2 being insignificant for $x = 1$ and only marginally significant for $x = 0$.

Order Effects: Recency and Primacy. The order effects model also improves fit relative to the base model, but not as much as the bias or priors models. For both $x = 1$ and $x = 0$, λ is significantly greater than 1, suggesting a recency effect, rather than primacy (consistent with uncertainty about whether the association might change during the process). Again, the level of association does not appear as a determinant of the order effect, with λ_2 not being statistically significant for either value of x.

Combined Model: Biased Learning, Priors, and Order. The most comprehensive model incorporates bias, priors, and order effects; see Equation (6). This has the best fit in Table 3 with respect to the Bayesian information criterion (BIC) for both $x = 1$ and $x = 0$. Bias is again greater than 1 but decreasing for $x = 1$ with a crossover point from overlearning to underlearning at $f_{1|1} = 0.81$, similar to that of the bias model. The values of prior means (μ) again fall close to 50/50. The strength of the prior beliefs (δ_1) is only marginally significant, and again, they are not contingent on the level of association δ_2. (If we omit δ_2

from the combined model, δ_1 becomes considerably more significant. However, we decided to report the complete model because the results are very similar.) Again, order effects are statistically significant, with λ_1 being significantly different from 1 for $x = 1$ and λ_2 being statistically significant for $x = 0$, driving λ beyond 1 for quite low levels of association.

In summary, in this context consumers are not best represented as statisticians (Model 1). Their learning about attribute–goal associations exhibits bias (Model 2), the effects of prior beliefs (Model 3), and order effects (Model 4). A combination of all three components appears to fit the data at least comparably to those with only one of the phenomena included. The trade-off between parsimony and improved fit appears to favor more complete models, particularly the incorporation of priors, which gives the greatest single reduction in log likelihood. We note that bias is the only component that is consistently related to the level of association (and this is for the case of $x = 1$).

5.2. Comparative Fit Relative to Benchmark Models

Whereas the previous results show that our models capture consumers' departures from rational learning in a simple and well-grounded way, other researchers have chosen different functional representations to describe some of these phenomena. We need to show that our models do not lead to a substantial loss of explanatory power relative to them.

For an alternative view of bias, we assume that rather than misclassifying a constant *proportion* of the successful goal outcomes ($\hat{\theta}^t_{1|1} = pf^t_{1|1}$), the subject misclassifies a constant *number* ($\hat{\theta}^t_{1|1} = P + f^t_{1|1}$). We also allow for bias to be a function of relative frequency in the alternative formulation: $P = P_1 + P_2 f^t_{1|1}$.

For the incorporation of priors, we use a partial adjustment model in which the subject updates her estimate each period by a proportion, D, of the difference between the information she is seeing and her previous estimate: $\hat{\theta}^t_{1|1} = f^t_{1|1} + D(\mu - f^t_{1|1})$. This latter model is popular in economics (e.g., Kennan 1979) and also in psychology in the form of error reduction models (Lopez et al. 1998a), though it lacks the simple beta-Bernoulli properties of our proposed model. Again, we allow for updating rate to be a function of the relative frequency: $D = D_1 + D_2 f^t_{1|1}$.

Finally, for an alternative way to represent order effects, we weight sample observations by the time since they were observed, $(t - i)$, raised to a constant power (e.g., Lovett 1998):

$$\hat{\theta}^t_{1|1} = \frac{\sum_{i=1}^{t} x_i y_i (t - i + 1)^L}{\sum_{i=1}^{t} x_i (t - i + 1)^L}.$$

We allow L to vary as a function of frequency: $L = L_1 + L_2 f^t_{1|1}$.

The estimation results for these alternative models are given in Table 4, along with the fit of our proposed models from Table 3 for comparison.

For the bias model, discounting by an absolute amount rather than a proportion of goal realizations seems to provide a slightly better representation of the data.[8] However, the difference is small, and we prefer the behavioral interpretation of our model. For the alternative models of priors and order effects, the fits are not as good as those of our proposed models according to the BIC criterion.

5.3. From Probabilities to Utilities

Having collected measures of both probabilities and utilities, we can also examine the probability to utility transformation. We estimate Equation (15), noting that the utility model proposed in Equation (12) predicts that $\tau = 0$ and $\rho = U_Y = 100$. In the second column of Table 5, we use the actual frequencies that subjects saw to predict their reported utilities. In column 3, we use their elicited probabilities, whereas in columns 4, 5, 6, and 7, we use the fitted probabilities from the bias, prior, order effects, and combined models, respectively. In all cases, the significance of σ^2_ν suggests that there are significant differences between subjects. Also, the significant negative coefficient of r_1 in all of the regressions suggests that the accuracy with which the subject translates her associations into a utility also improves over replications. For most models, the intercept τ is not significantly different from 0. Although in none of the cases do subjects totally value the attribute as much as the probability or their estimate of the probability suggests that they should (ρ is significantly less than 100), it is interesting that this effect is least for our two models that fit the consumers' association evolution using bias and priors.[9] Whereas reported probabilities explain perceived utilities better than estimated ones, three of our models of associations provide a better fit to utility than actual associations, suggesting some internal consistency to our approach.

5.4. Predictive Ability and Parameter Stability

The final test of our models is to examine their predictive ability. Our approach may be useful to managers

[8] This is not surprising because the inclusion of P_2 indirectly provides a form of proportional bias similar to p_1 in our proposed model and thus subsumes it. Note that the P_2 component is considerably more significant than P_1.

[9] There are a number of reasons why utility as measured by willingness to pay might be somewhat lower than the expected utility calculated from $(\theta_{1|1} - \theta_{1|0})U_Y$, the major one being risk aversion. These differences are surprisingly small except for the elicited probabilities case (column 3), which appears to be an anomaly.

Table 4 Benchmarking Proposed Models: Alternative Models of Association for $\theta_{1|0}$ and $\theta_{1|1}$ (Noise Data)

	Bias (Absolute)		Priors (Partial adjustment)		Order effects (Lovett)	
Parameter	$\theta_{1\|1}$	$\theta_{1\|0}$	$\theta_{1\|1}$	$\theta_{1\|0}$	$\theta_{1\|1}$	$\theta_{1\|0}$
σ_v^2	0.01***	0.01***	0.01***	0.01***	0.01***	0.07***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
c_0	5.53***	5.27***	5.54***	5.31***	6.15***	5.63***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
c_1	−0.94***	−0.36***	−0.91***	−0.44***	−1.25***	0.25***
	(0.10)	(0.14)	(0.10)	(0.14)	(0.14)	(0.09)
r_0	6.65***	6.61***	6.80***	6.79***	5.81***	3.37**
	(0.18)	(0.18)	(0.17)	(0.18)	(0.18)	(1.29)
r_1	−0.60***	−0.63***	−0.66***	−0.69***	−0.31***	−0.21
	(0.09)	(0.10)	(0.09)	(0.11)	(0.10)	(0.34)
P_1	0.03**	0.10***				
	(0.01)	(0.01)				
P_2	−0.36***	−0.53***				
	(0.06)	(0.08)				
μ			0.66***	0.98***		
			(0.10)	(0.11)		
D_1			0.04	0.09***		
			(0.05)	(0.03)		
D_2			0.14	−0.28***		
			(0.10)	(0.09)		
L_1					0.64***	0.15*
					(0.01)	(0.09)
L_2					−0.10***	−0.17*
					(0.01)	(0.09)
LL	−13,378	−13,506	−13,389	−13,529	−13,937	−14,200
BIC	26,790	27,045	26,816	27,096	27,908	28,434
Fit of proposed models (from Table 3)						
LL	−13,387	−13,536	−13,357	−13,487	−13,395	−13,517
BIC	26,808	27,107	26,752	27,013	26,825	27,069

Notes. Standard errors are in parentheses. The number of observations for all models is 3,240.
*Denotes significance at the 10% level; **denotes significance at the 5% level; and ***denotes significance at the 1% level.

Table 5 Probability to Utility Transformation (Noise Data)

	Source of contingency (association) estimate					
Parameter	Actual data	Elicited probabilities	Model 2 (Bias)	Model 3 (Priors)	Model 4 (Order effects)	Combined (Bias + Priors + Order effects)
σ_v^2	418.39***	1,066.51***	402.52***	399.82***	419.48***	416.35***
	(53.73)	(142.60)	(51.86)	(51.40)	(55.85)	(55.08)
r_0	7.29***	7.23***	7.29***	7.20***	7.30***	7.19***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
r_1	−0.14***	−0.14***	−0.14***	−0.13***	−0.14***	−0.12***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
τ	−0.38	7.69***	−0.27	−0.42	0.95	0.70
	(1.89)	(2.86)	(1.85)	(1.84)	(1.88)	(1.87)
ρ	85.70***	41.90***	90.12***	92.26**	82.84***	83.82***
	(3.09)	(1.99)	(3.19)	(3.07)	(3.19)	(3.01)
LL	−15,164	−15,113	−15,161	−15,116	−15,184	−15,140
BIC	30,351	30,251	30,346	30,255	30,392	30,304

Notes. Standard errors are in parentheses. The significance of ρ is tested relative to 100. The number of observations for all models is 3,240.
*Denotes significance at the 10% level; **denotes significance at the 5% level; and ***denotes significance at the 1% level.

in two major ways. First, if the manager understands the drivers of importance weight formation (i.e., what influences the rate at which it progresses and the level to which it asymptotes), she may direct her marketing actions to accelerate and accentuate it. Second, an understanding of attribute utility evolution may be useful in predicting the ultimate demand for a product for planning purposes. We do not manipulate the learning environment, so we cannot test different managerial strategies. However, we can test our models' predictive accuracy. To ascertain our models' predictive accuracy, we truncate our data after the ninth replication, reestimate the models, and test their predictions of subjects' reports for replications 10, 11, and 12, giving us one-, two-, and three-step-ahead forecasts. This approach provides the added benefit of testing the stability of our approach over different estimation periods.

The results of fitting replications 1 to 9 are very similar to those for replications 1 to 12 in Table 3. They are presented in Table B1 in Web Appendix B. For example, just looking at the combined model, individual response variance σ_v^2 is again significant, as are the frequency and replication effects on the variance. For $x = 1$, the subject again starts with overlearning but moves to underlearning at higher frequencies (with the crossover point being $f_{1|1} = 0.88$). Again, bias is not so strong in the absence of the attribute. For $x = 1$, prior and order (recency) effects are again significant but are not frequency dependent.

In terms of forecasting ability, the one-, two-, and three-step-ahead average forecast errors using the combined model are 7.92, 9.46, and 9.25 for $x = 1$, respectively, and 8.30, 9.49, and 10.20 for $x = 0$. (By way of reference, remember that we are modeling associations that range from 0 to 100. The average absolute errors for subjects' self-stated associations are 9.58 for $x = 1$ and 10.82 for $x = 0$. Thus, the model leads to an average of a 10.55% decrease in forecast error.[10])

6. Illustrative Extensions

The modeling and calibration framework that we developed in §§3 and 4, and tested in §5, lends itself to a number of extensions. We present two of these to illustrate how the framework might be used in alternative contexts. The first relates to the data generation process. We examine whether learning occurs differently if there is no noise in the observations subjects see. The second relaxes the assumption that there is only one diagnostic attribute about which the consumer can learn. We look specifically at the two-attribute case.

6.1. Estimation of the Models Using a Stimulus with No Noise

In the second experiment, all goal realizations were generated at the mean association, rather than drawn from a distribution. Experiment 2 enables us to compare the results between subjects learning in an environment with no noise to those who learn with noise. For example, Kahn and Meyer (1991) suggest that random noise obscures the consistency of the underlying stimulus, meaning that in unfamiliar categories, one might see greater departures from rational decision making. Most experiments in psychology present associations to subjects at their expected level of occurrence. That is, if the association of a stimulus and an outcome is 0.75, subjects always see three outcomes for every four occurrences of the stimulus. The studies by Lopez et al. (1998a, b) are but two of the many examples. This approach is inconsistent with most assumed data-generating processes where a subject is considered to take draws from a distribution. If the expected co-occurrence between the presence of an attribute and the achievement of a goal is 0.75, then according to the Bernoulli distribution in four sample realizations, 32% of the time we would except to see four co-occurrences, while 0.4% of the time, we would expect to see none. In the no-noise experiment, subjects always saw three co-occurrences (in keeping with traditional approaches), in contrast to the noise experiment where we used the association to generate sample realizations drawing from a Bernoulli distribution.

The ranges of associations studied and results for the no-noise experiment are given in Tables C1–C3 in Web Appendix C. In summary, the fits of the various no-noise models for associations are remarkably similar to those for the noise condition (as given in Table 3), which is a source of comfort for traditional approaches. (Note that it is not possible to estimate the order model on the no-noise data because there is no difference between the stimuli the subject sees in each replication of the experiment.) The parameters are almost identical and the only major difference (unsurprisingly) is the degree of fit, which is better for the no-noise experiment. As with the noise experiment, random effects and contingency- and replication-specific variances are all significant. Again for $x = 1$, the subject goes from overlearning to underlearning with respect to bias. Priors effects are in evidence for $x = 1$ but are not significantly related to association levels. Again, incoming priors are relatively close to 0.5. The results for the association to utility transformation are also similar to those in Table 5 with the exception that with no noise, the

[10] The forecasts of Models 2, 3, and 4 also outperform the average absolute error, with the bias and priors models predicting slightly better than the combined model and the order model predicting slightly worse.

combined and prior models are not statistically different from that predicted by Equation (12) (i.e., there is no undervaluation of attributes). It would therefore appear that when noise is added to the signal of a product attribute's efficacy in achieving a goal, consumers make larger errors in determining associations (but they still learn to the same degree), and they do not make the association to utility link perfectly.

6.2. Extension to the Two-Attribute Case

To date, we have studied only one active attribute because the separability assumption suggests that it is often possible to examine consumer learning of attribute importances one at a time. There are two reasons why this might not be a good assumption. First, even if the effect of any attribute on the attainment of a goal is independent of the level of others, the cognitive complexity of learning about more than any one attribute at a time may interfere with the learning of others. For example, an attribute that has a high contingency with goal attainment may mask the effect of, and decrease the incentive to learn about, a less effective attribute. This phenomenon is well documented in psychology in terms of cue interference, including overshadowing (where a strongly salient attribute masks learning about a weakly salient one; e.g., see Price and Yates 1993) and blocking (where initial learning about one attribute goal association prevents learning about the association of another; e.g., see Chapman and Robbins 1990).

Second, the effects of different attributes may not be independent. For example, Shanks et al. (1998) examine the case of configural learning, where participants learn about the association of a combination of attributes with a goal but do not learn about individual attributes on their own. Most modeling applications in marketing assume independence of the effect of different attributes on the utility function. However, the case of correlated effects is an important one. Kayande et al. (2007) show how correlated attribute effects may influence not only the expected value obtained but also the certainty with which the consumer evaluates a product.

In this section we extend our study in an illustrative way to the case of two attributes, x and z, both of which are diagnostic in determining the utility of the product. (Remember that in the main study we did have two attributes, but the level of the second attribute did not influence the probability of goal achievement.) We consider both uncorrelated- and correlated-attribute cases.

Modeling Utility with More Than One Attribute. In the case of two binary attributes, x and z, the consumer is faced with four possible product combinations. If we denote the probability of goal attainment or not in the presence of x and z by $\theta_{y|xz}$, then to estimate the utility of a product with different attribute combinations, our framework suggests that the consumer must form estimates of $\theta_{1|11}$, $\theta_{1|10}$, $\theta_{1|01}$, and $\theta_{1|00}$ (where, for example, $\theta_{1|01}$ represents the probability of goal achievement when $x = 0$ and $z = 1$). If we assume that the consumer is a cognitive miser (Fiske and Taylor 1991) and she also understands the value of an attribute in terms of the extent to which it helps her in achieving a salient goal times the value of that goal (Equation (12)), then the easiest way for her to assess the value of an attribute is to first estimate the four conditional probabilities above separately. In the multiattribute case, learning about associations may be slower, may lead to increased bias, or may cause order effects to be more pronounced than single-attribute learning, but we would expect Equations (1), (2), (3), (5), and (6) to still pertain.

We note that when $\theta_{1|11} = \theta_{1|10}$ and $\theta_{1|01} = \theta_{1|00}$, the level of z does not affect goal attainment. Attribute z is diagnostic, but its effect is independent of that of x when $\theta_{1|11} - \theta_{1|01} = \theta_{1|10} - \theta_{1|00}$ and $\theta_{1|11} \neq \theta_{1|10}$. Dependence is said to exist when $\theta_{1|11} - \theta_{1|01} \neq \theta_{1|10} - \theta_{1|00}$.

In terms of the association to utility transformation, the separable multiattribute utility Equation (13) applies in the case of independent attributes. In the case of correlated effects, we would expect to see an interaction effect. By substituting different values for x and z in the expression for expected utility, we may readily see that the generalized two-attribute version of Equation (13) can be written as follows:

$$EU(x, z) = \gamma_0 + \gamma_x x + \gamma_z z + \gamma_{xz} xz, \qquad (16)$$

where $\gamma_0 = \theta_{1|00} U_Y$, $\gamma_x = (\theta_{1|10} - \theta_{1|00}) U_Y$, $\gamma_z = (\theta_{1|01} - \theta_{1|00}) U_Y$, and $\gamma_{xz} = (\theta_{1|11} - \theta_{1|10} - \theta_{1|01} + \theta_{1|00}) U_Y$. We note from the above discussion that when z is diagnostic but its effect is independent of x, $\gamma_{xz} = 0$, and therefore Equation (16) collapses to the separable utility Equation (13).

Testing the Two-Attribute Model. We tested the two-attribute model in an analogous setting to that described in §4. We used the same product and same two attributes (Neural Net and Temporally Continuous for x and z, respectively), except in this case, we allowed the latter to also affect the probability of goal achievement. We tested four conditions, illustrated in Table 6. Condition 1 acts as a control to establish comparability with the main experiment. The level of z does not affect goal attainment, and thus it is equivalent to condition 6 in Table 2. In conditions 2 and 3, z is diagnostic (i.e., it does influence the probability of goal achievement), but its effect is independent of that of x. In condition 2, x has a high contingency

Table 6 Conditional Probabilities and Realized Associations in Experiment 3: Multiple Attributes (Noise Design)

	Programmed conditional probabilities				Realized conditional probabilities			
Condition	$\theta_{1\|11}$	$\theta_{1\|10}$	$\theta_{1\|01}$	$\theta_{1\|00}$	$\theta_{1\|11}$	$\theta_{1\|10}$	$\theta_{1\|01}$	$\theta_{1\|00}$
1 Control	0.75	0.75	0.25	0.25	0.73	0.79	0.19	0.21
2 Uncorrelated (high)	1.00	0.75	0.50	0.25	1.00	0.82	0.50	0.23
3 Uncorrelated (low)	0.75	0.50	0.50	0.25	0.76	0.48	0.49	0.23
4 Correlated	0.75	0.75	0.50	0.25	0.75	0.77	0.53	0.25

$(\theta_{1|10} - \theta_{1|00} = 0.5)$, while in condition 3, it is moderate $(\theta_{1|10} - \theta_{1|00} = 0.25)$. In condition 4, the independence assumption is relaxed $(\theta_{1|11} - \theta_{1|01} = 0.25 \neq \theta_{1|10} - \theta_{1|00} = 0.5)$. We had 30 subjects for the control condition and 45 for each of the remaining three test conditions.

Results of Fitting the Two-Attribute Model. We examine the results of fitting the two-attribute model in two stages. First, we examine the learning of associations to see whether the rate or level of associative learning is different in the cases of a diagnostic uncorrelated or correlated second attribute. In view of its strong performance in §5, we only fit the combined model.

After fitting the associative learning model, we proceed to examine the association to utility transformation to see whether its nature is affected by the presence of the second attribute. We are unable to pool across associations, as was the case in §5, and so we have considerably fewer degrees of freedom. To increase our statistical power, we pooled the four associations in each condition $(\theta_{1|11}, \theta_{1|10}, \theta_{1|01}, \theta_{1|00})$, assuming they share the same learning parameters.

Table 7 presents the two-attribute analog to Table 3. Association-specific parameters (c_1, c_2, p_2, δ_2, and λ_2) are not relevant because we do not estimate across associations. Random effects and replication-specific variables are again statistically significant. In the high-association conditions (control, high, and correlated), we see underlearning as expected, whereas in the low-association condition, we do not. The effect sizes are similar to those of learning with one attribute. However, whereas priors and their updating are still significant, the amount of information needed to overcome priors in the more complex information environment is higher (δ is larger). Order effects are only significant for the uncorrelated high-association condition, a fact that may be due to the decreased statistical power of our test here.

The association to utility transformation calibration results are presented in Table 8, and it is here that the surprise really comes. Whereas random effects and replication variances are consistent with those in Table 5, estimates of the intercept (τ) and slope parameters of contingency (ρ) are considerably changed. The intercept is now highly significant and the effect of contingency on utility is strongly diminished from its rational value of 100. These multiattribute examples are positioned as illustrative because of the wide range of associations and correlations that could have been investigated, but they do seem to provide strong circumstantial evidence that with more than one diagnostic attribute, the consumer has less trouble in determining whether an attribute is useful in accomplishing her goal and more difficulty in deciding how to value the attribute based on that association. If that is the case more broadly, this has obvious implications for communications strategies. A brand with one new attribute may be able to establish its efficacy and hence utility based on these results. Where two new attributes offer increased utility, the manager's communication problem may be less about establishing the usefulness of the attributes

Table 7 Estimates of $\theta_{1|0}$ and $\theta_{1|1}$ for Proposed Models of Association (Two-Active-Attribute Case)

Parameter	Control	Uncorrelated (high)	Uncorrelated (low)	Correlated
σ_v^2	0.01***	0.00***	0.00***	0.01***
	(0.00)	(0.00)	(0.00)	0.00
r_0	6.45***	6.60***	6.49***	6.40***
	(0.08)	(0.07)	(0.07)	0.07
r_1	−0.02	−0.06***	−0.03***	−0.01
	(0.01)	(0.01)	(0.01)	0.01
p	0.88***	0.94***	1.03	0.92***
	(0.04)	(0.02)	(0.03)	0.03
μ	0.55***	0.49***	0.48***	0.54***
	(0.03)	(0.03)	(0.02)	0.02
δ	11.23***	9.91***	18.07***	14.07***
	(2.40)	(1.37)	(2.65)	2.05
λ	0.95	1.15***	1.03	1.03
	(0.06)	(0.04)	(0.03)	0.03
LL	−6,632	−9,831	−9,877	−9,940
n	1,440	2,160	2,160	2,160
BIC	13,288	19,688	19,781	19,906

Notes. Standard errors are in parentheses. The significance of p and λ are tested relative to 1.0.

*Denotes significance at the 10% level; **denotes significance at the 5% level; and ***denotes significance at the 1% level.

Table 8 Probability to Utility Transformation (Two-Attribute Case)

	Control			Uncorrelated (high)		
Parameter	Actual data	Elicited probabilities	Model	Actual data	Elicited probabilities	Model
σ_v^2	204.67***	344.39*	162.22***	81.76***	155.99***	107.85***
	(56.79)	(174.59)	(37.52)	(19.45)	(34.96)	(32.57)
r_0	6.68***	6.45***	6.68***	6.76***	6.56***	6.78***
	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)
r_1	−0.01	−0.00	−0.01	−0.04***	−0.07***	−0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
τ	31.32***	30.46***	31.19***	29.94***	24.61***	31.69***
	(2.77)	(3.47)	(2.51)	(1.82)	(1.99)	(1.94)
ρ	26.16***	42.14***	47.18***	34.48***	54.01***	38.17***
	(2.12)	(2.03)	(3.91)	(2.99)	(1.61)	(3.57)
LL	−6,836	−6,723	−6,839	−10,100	−9,722	−10,110
n	1,440	1,440	1,440	2,160	2,160	2,160
BIC	13,688	13,462	13,695	20,219	19,462	20,239
	Uncorrelated (low)			**Correlated**		
Parameter	Actual data	Elicited probabilities	Model	Actual data	Elicited probabilities	Model
σ_v^2	201.39***	205.51***	200.72***	202.91***	231.19**	203.85***
	(65.63)	(59.09)	(65.23)	(55.26)	(90.16)	(55.05)
r_0	6.44***	6.29***	6.44***	6.53***	6.44***	6.55***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
r_1	−0.02**	−0.04***	−0.02**	−0.02**	−0.03***	−0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
τ	33.77***	25.70***	34.00***	30.58***	28.76***	30.79***
	(2.42)	(2.21)	(2.44)	(2.26)	(2.34)	(2.27)
ρ	−0.21***	39.85***	−1.82***	21.11***	37.39***	33.89***
	(3.84)	(1.58)	(6.86)	(2.17)	(1.62)	(3.75)
LL	−9,923	−9,657	−9,923	−10,041	−9,852	−10,048
n	2,160	2,160	2,160	2,160	2,160	2,160
BIC	19,864	19,332	19,864	20,100	19,722	20,114

Notes. Standard errors are in parentheses. The significance of ρ is tested relative to 100.
*Denotes significance at the 10% level; **denotes significance at the 5% level; and ***denotes significance at the 1% level.

in achieving a goal and more about showing the consumer why that usefulness should lead to a higher valuation. This is consistent with popular advertising lore, which suggests that the advertiser should aim to own a single attribute in the customer's mind (e.g., Ries and Trout 1994).

7. Discussion and Conclusion

In this paper we developed and tested a set of models of binary attribute importance weight formation. We interpreted attribute importance in terms of an associative relationship between a consumer's goal and a product attribute, and we modeled the learning of these relations based on the classical conditioning process. Our models of importance weight evolution, motivated by Lancaster (1966), allow us to separate attribute importance into properties intrinsic to consumer's goal and those relating to the attribute's incremental ability to achieve that goal. In our application, we control the subjects' goal and allow them to learn the attribute's ability to satisfy it.

We tested for the existence of different learning phenomena using our models. We find that consumers may have prior expectations about the importance of unfamiliar product attributes in a new product category but will revise them in the face of evidence, though often to an equilibrium position not exactly equal to that of a rational decision maker. We also observe overall recency effects in our particular application. Our representation is parsimonious, is linked directly to multiattribute utility theory, and fits at least comparably to alternative formulations previously proposed to (separately) look at other forms of learning. We find that an attribute's importance evolution depends on the association between its presence and the achievement of a goal. Similar results pertain in environments in which there is no noise (and, not surprisingly, these models tend to fit the data better than in the case of noise). We also illustrated our models in the case of learning about more than one active attribute. We found similar results for associative learning but much weaker evidence of that association being translated into utility judgments. In other applications the relative importance of these phenomena may be different.

The implications for researchers of having a methodology to determine importance weights over time is that it enables them to understand one important aspect of preference and choice dynamics for products with new attributes. For managers, our approach provides a method of calibrating likely equilibrium attribute importance and how that importance will evolve once consumers' goals have been uncovered using market research. It will be useful for prelaunch studies where it is realistic to expose consumers to multiple realizations of their goal outcomes in the presence of the product. However, it may also be used postlaunch to track attribute valuation based on consumers' experience with the product. It could also be combined with choice-based conjoint analysis to study one aspect of the dynamics of choice.[11]

Our research does make assumptions that bring with them a number of limitations. Our main experiment relied on a single, stable, and well-defined consumer goal, and all three experiments used binary product attributes in a new product category. This scenario is useful for demonstrating the key predictions of the models and examining the role of different levels of association. Moreover, the case of binary attributes is prevalent in many new product markets. The generalizability of our approach to situations involving multilevel predictor attributes, as well as competing and/or fuzzy consumer goals, can be undertaken with further research. For example, the importance weights of continuous attributes could be represented by the effect of sample realizations of different levels and the achievement (or partial achievement) of goals using continuous Bayesian conjugate pairs. Phenomena beyond bias, priors, and order effects could also be incorporated into our models. Also, it would be useful to study how consumer learning varies as associations change (for example, an attribute becomes more efficacious as a result of quality improvements) or when there is serial autocorrelation (successes and failures tend to come in runs).

A natural extension of the research would be to investigate the effect of different interventions on learning so that the models could be used not only to estimate importance weight evolution but also to test different management strategies to influence it. As a first step, what we have done is to develop and apply a simple set of utility learning models that captures a number of key phenomena in binary importance weight evolution and which allows us to assess the presence of different phenomena, test their forecasting ability, and see how they vary with association level under two data-generating processes.

[11] We are grateful to an anonymous reviewer for this suggestion.

Electronic Companion

An electronic companion to this paper is available as part of the online version at http://dx.doi.org/10.1287/mksc.1120.0719.

Acknowledgments

The authors thank the editor, Preyas Desai, the associate editor, and the reviewers for their support and advice throughout the review process. They also thank the Australian School of Business for the use of its ASBLab research facility, and acknowledge suggestions made by participants at the Waiheke Marketing Conference. The third author thanks the Australian National University for support during his stay in Canberra while working on this research.

References

Alba JW, Hutchinson JW (2000) Knowledge calibration: What consumers know and what they think they know. *J. Consumer Res.* 27(2):123–156.

Alloy LB, Tabachnik N (1984) Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psych. Rev.* 91(1):112–149.

Anderson JR, Lebiere C (1998) *The Atomic Components of Thought* (Lawrence Erlbaum Associates, Mahwah, NJ).

Bass FM, Bruce N, Majumdar S, Murthi BPS (2007) Wearout effects of different advertising themes: A dynamic Bayesian model of advertising-sales relationship. *Marketing Sci.* 26(2):179–195.

Bettman JR, Luce MF, Payne JW (1998) Constructive consumer choice processes. *J. Consumer Res.* 25(3):187–217.

Blackorby C, Primont D, Russell RR (1977) On testing separability restrictions with flexible functional forms. *J. Econometrics* 5(2):195–209.

Bradlow ET, Hu Y, Ho T-H (2004) Modeling behavioral regularities of consumer learning in conjoint analysis. *J. Marketing Res.* 41(4):392–396.

Cason T, Friedman D (1999) Learning in a laboratory market with random supply and demand. *Experiment. Econom.* 2(1):77–98.

Chapman GB, Robbins SJ (1990) Cue interaction in human contingency judgment. *Memory Cognition* 18(5):537–545.

Cheng PW (1997) From covariation to causation: A causal power theory. *Psych. Rev.* 104(2):367–405.

Fader PS, McAlister L (1990) An elimination by aspects model of consumer response to promotion calibrated on UPC scanner data. *J. Marketing Res.* 27(3):322–332.

Fiske ST, Taylor SE (1991) *Social Cognition* (McGraw-Hill, New York).

Gerber A, Green D (1999) Misperceptions about perceptual bias. *Annual Rev. Political Sci.* 2:189–210.

Gönül F, Srinivasan K (1996) Estimating the impact of consumer expectations of coupons on purchase behavior: A dynamic structural model. *Marketing Sci.* 15(3):262–282.

Griffiths TL, Tenenbaum JB (2005) Structure and strength in causal induction. *Cognitive Psych.* 51(4):354–384.

Guadagni PM, Little JDC (1983) A logit model of brand choice calibrated on scanner data. *Marketing Sci.* 2(3):203–238.

Hoch SJ, Deighton J (1989) Managing what consumers learn from experience. *J. Marketing* 53(2):1–20.

Hoeffler S, Ariely D (1999) Constructing stable preferences: A look into dimensions of experience and their impact on preference stability. *J. Consumer Psych.* 8(2):113–139.

Hoeffler S, Ariely D, West P (2006) Path dependent preferences: The role of early experience and biased search in preference development. *Organ. Behav. Human Decision Processes* 101(2):215–229.

Iyengar R, Ansari A, Gupta S (2007) A model of consumer learning for service quality and usage. *J. Marketing Res.* 44(4):529–544.

Janiszewski C, van Osselaer SMJ (2000) A connectionist model of brand-quality associations. *J. Marketing Res.* 37(3):331–350.

Janiszewski C, Warlop L (1993) The influence of classical conditioning procedures on subsequent attention to the conditioned brand. *J. Consumer Res.* 20(2):171–189.

Jedidi K, Mela CF, Gupta S (1999) Managing advertising and promotion for long-run profitability. *Marketing Sci.* 18(1):1–22.

Johar GV, Jedidi K, Jacoby J (1997) A varying parameter averaging model of on-line brand evaluations. *J. Consumer Res.* 24(2):232–248.

Kahn BE, Meyer RJ (1991) Consumer multiattribute judgments under attribute-weight uncertainty. *J. Consumer Res.* 17(4): 508–522.

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decisions under risk. *Econometrica* 47(2):263–292.

Kardes FR, Posavac SS, Cronley ML (2004) Consumer inference: A review of processes, bases, and judgment contexts. *J. Consumer Psych.* 14(3):230–256.

Kayande U, Roberts JH, Lilien GL, Fong D (2007) Mapping the bounds of incoherence: How far can you go and how does it affect your brand? *Marketing Sci.* 26(4):508–514.

Keeney RL, Raiffa H (1976) *Decisions with Multiple Objectives: Preferences and Value Tradeoffs* (John Wiley & Sons, New York).

Kennan J (1979) The estimation of partial adjustment models with rational expectations. *Econometrica* 47(6):1441–1455.

Kruschke JK (2006) Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psych. Rev.* 113(4):677–699.

Kuehn AA (1958) An analysis of the dynamics of consumer behavior and its implications for marketing management. Unpublished doctoral dissertation, Carnegie Institute of Technology, Carnegie Mellon University, Pittsburgh.

Lancaster KJ (1966) A new approach to consumer theory. *J. Political Econom.* 74(2):132–157.

Lawson R (1997) Consumer decision making within a goal-driven framework. *Psych. Marketing* 14(5):427–449.

Lilien GL (1974) A modified linear learning model of buyer behavior. *Management Sci.* 20(7):1027–1036.

Lopez FJ, Cobos PL, Cano A, Shanks DR (1998a) The rational analysis of human and animal probability judgments. Oaksford M, Chater N, eds. *Rational Models of Cognition* (Oxford University Press, Oxford, UK), 314–352.

Lopez FJ, Shanks DR, Almaraz J, Fernandez P (1998b) Effects of trial order on association judgments: A comparison of associative and probabilistic contrast accounts. *J. Experiment. Psych.: Learn., Memory, Cognition* 24(3):627–694.

Lovett M (1998) Choice: Elaborating ACT-R's theory of choice. Anderson JR, Lebiere C, eds. *The Atomic Components of Thought* (Lawrence Erlbaum Associates, Mahwah, NJ), 255–296.

Marsh JK, Ahn W (2006) Order effects in association learning: The role of task complexity. *Memory Cognition* 34(3):568–576.

Massy WF, Montgomery DB, Morrison DG (1970) *Stochastic Models of Buying Behavior* (MIT Press, Cambridge, MA).

McFadden D (2001) Economic choices. *Amer. Econom. Rev.* 91(3):351–379.

McGough B (2003) Statistical learning with time-varying parameters. *Macroeconomic Dynam.* 7(1):119–139.

Miller RR, Matzel LD (1988) The comparator hypothesis: A response rule for the expression of associations. Bower GH, ed. *The Psychology of Learning and Motivation*, Vol. 22 (Academic Press, San Diego), 51–92.

Pearce JM (1994) Similarity and discrimination: A selective review and a connectionist model. *Psych. Rev.* 101(4):587–607.

Pleskac TJ (2008) Decision making and learning while taking sequential risks. *J. Experiment. Psych.: Learn., Memory, Cognition* 34(1):167–185.

Price PC, Yates JF (1993) Judgmental overshadowing: Further evidence of cue interaction in contingency judgment. *Memory Cognition* 21(5):561–572.

Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Black AH, Prokasy WF, eds. *Classical Conditioning II* (Appleton-Century-Crofts, New York), 64–99.

Ries A, Trout J (1994) *The 22 Immutable Laws of Marketing* (HarperCollins, London).

Roberts JH, Urban GL (1988) Modeling multi-attribute utility, risk, and belief dynamics for new consumer durable brand choice. *Management Sci.* 34(2):167–185.

Schooler LJ, Raaijmakers JGW, Shiffrin RM (2001) A Bayesian model of implicit effects in perceptual identification. *Psych. Rev.* 108(1):257–272.

Shanks DR, Charles D, Darby RJ, Azmi A (1998) Configural processes in human associative learning. *J. Experiment. Psych.: Learn., Memory, Cognition* 24(6):1353–1378.

Simonson I, Tversky A (1992) Choice in context: Tradeoff contrast and extremeness aversion. *J. Marketing Res.* 29(3):281–295.

Sood A, Tellis GJ (2005) Technological evolution and radical innovation. *J. Marketing* 69(3):152–168.

van Osselaer SMJ, Alba JW (2000) Consumer learning and brand equity. *J. Consumer Res.* 27(1):1–16.

van Osselaer SMJ, Alba JW (2003) Locus of equity and brand extension. *J. Consumer Res.* 29(4):539–550.

van Osselaer SMJ, Janiszewski C (2001) Two ways of learning brand associations. *J. Consumer Res.* 28(2):202–223.

Wallsten T, Pleskac T, Lejuez C (2005) Modeling behavior in a clinically diagnostic sequential risk-taking task. *Psych. Rev.* 112(4):862–880.

Wasserman EA, Miller RR (1997) What's elementary about associative learning? *Annual Rev. Psych.* 48:573–607.

Xiao B, Benbasat I (2007) E-commerce product recommendations: Use, characteristics, and impact. *MIS Quart.* 31(1):137–209.