

Predictably Non-Bayesian: Quantifying Salience Effects in Physician Learning About Drug Quality

Nuno Camacho, Bas Donkers

Erasmus School of Economics, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands
{camacho@ese.eur.nl, donkers@ese.eur.nl}

Stefan Stremersch

Erasmus School of Economics, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands;
and IESE Business School, University of Navarra, 08034 Barcelona, Spain, stremersch@ese.eur.nl

Experimental and survey-based research suggests that consumers often rely on their intuition and cognitive shortcuts to make decisions. Intuition and cognitive shortcuts can lead to suboptimal decisions and, especially in high-stakes decisions, to legitimate welfare concerns. In this paper, we propose an extension of a Bayesian learning model that allows us to quantify the impact of salience—the fact that some pieces of information are easier to retrieve from memory than others—on physician learning. We show, using data on actual prescriptions for real patients, that physicians’ belief formation is strongly influenced by salience effects. Feedback from switching patients—the ones the physician decided to switch to a clinically equivalent treatment—receives considerably more weight than feedback from other patients. In the category we study, salience effects slowed down physicians’ speed of learning and the adoption of a new treatment, which raises welfare concerns. For managers, our findings suggest that firms that are able to eliminate, or at least reduce, salience effects to a greater extent than their competitors can speed up the adoption of new treatments. We explore the implications of these results and suggest alternative applications of our model that are relevant for policy makers and managers.

Key words: consumer learning; quasi-Bayesian learning models; behavioral modeling; medical decision making; physician learning; new drug adoption

History: Received: October 23, 2008; accepted: November 5, 2010; Eric Bradlow served as the editor-in-chief and Teck Ho served as associate editor for this article. Published online in *Articles in Advance* January 21, 2011.

1. Introduction

Scholars in marketing and economics have developed Bayesian updating models for consumer (e.g., Erdem and Keane 1996, Mehta et al. 2008, Roberts and Urban 1988) and physician learning (Coscelli and Shum 2004, Crawford and Shum 2005, Narayanan et al. 2005, Narayanan and Manchanda 2009). Bayesian learning enables researchers to structurally model the evolution of an agent’s belief about any uncertain attribute, e.g., about the quality of a product, by integrating new information and prior beliefs using Bayes’ rule. Bayes’ rule is the normative way to update probabilistic beliefs; i.e., these models assume that decision makers learn using an optimal rule. However, many scholars claim that the assumptions behind Bayesian learning are not psychologically or cognitively valid (see, e.g., Camerer and Loewenstein 2004).

In particular, consumers often deviate from Bayes’ rule by giving more weight to more easily accessible, i.e., more salient, pieces of information they retrieve from memory. To model salience effects, we propose a

quasi-Bayesian learning model. Quasi-Bayesian learning models apply Bayes’ rule to subjectively revised evidence or prior beliefs (Epstein 2006, Rabin and Schrag 1999) and may “become the standard way for translating the cognitive psychology of judgment into a tractable alternative to Bayes’ rule” (Camerer and Loewenstein 2004, p. 13).

However, identification and estimation of quasi-Bayesian models is often difficult, and empirical applications using revealed preference data are still rare (for notable exceptions, see Mehta et al. 2004, 2008). In this paper, we study physician learning regarding the quality of a new treatment, a context where consumers are particularly sophisticated and involved in the choice process, as the stakes are high. Our proposed model fits in the rapidly growing field of behavioral modeling, a field that seeks to enrich mathematical models of consumer behavior, which are typically normative models, with robust psychological regularities (Häubl et al. 2010, Ho et al. 2006, Narasimhan et al. 2005).

Despite their specialized training, physicians often rely on their intuition and are selective in the use

of new information, deviating from normative rules in predictable ways, very much like humans in general (e.g., Croskerry 2002, Elstein and Schwartz 2002, Redelmeier 2005). However, the evidence accumulated about physicians' deviation from optimal reasoning and decision making thus far relies solely on experimental and survey research with physicians (Bornstein et al. 1999, Estrada et al. 1997, Poses and Anthony 1991) and participants role-playing as physicians (Medin et al. 1982) rather than research on actual physician decisions for real patients, as presented in the present paper.

We calibrate our model on a unique panel data set of Dutch general practitioner prescription behavior in the obstructive airways diseases category (i.e., treatments for asthma and chronic obstructive pulmonary disease). The data were retrieved from the Integrated Primary Care Information database, which is maintained by the School of Medicine of the Erasmus University Rotterdam (for a detailed description, see Vlug et al. 1999). These data are particularly well suited to test whether salience interferes with physicians' formation of treatment quality beliefs and if yes, to what extent. First, physicians in our data set use *paperless offices*, guaranteeing that the full clinical history of their patients gets stored in the database, which allows us to model treatment¹ choices using both new prescriptions and repeat prescriptions. Second, at the start of our observation period, a new treatment—AstraZeneca's Symbicort—was introduced in the category we study, which facilitates identification of dynamics in physicians' quality beliefs.

Our central hypothesis is that patients who the physician switches away from a specific treatment to a clinically equivalent alternative² become salient in the physician's memory. Consider the case of Dr. Jones, an imaginary general practitioner who sees about five or six patients with asthma complaints per week. Dr. Jones decides to prescribe a new brand—Symbicort—to 20 patients, i.e., about half of the asthma patients he sees in the first eight weeks after the launch of Symbicort. Two of these patients, Mrs. Smith and Mr. Miller, later complain that Symbicort made them dizzy, nauseated, and tired. To avoid future complaints, Dr. Jones switches Mrs. Smith and Mr. Miller to an older treatment alternative.

¹ We use the term *treatment* instead of *drug* because our empirical application is focused on treatments with two molecules, a preventive (anti-inflammatory) and a reliever (bronchodilator), either prescribed in two distinct inhalers or combined in a single-inhaler device.

² Clinically equivalent alternatives are those that can be considered as substitutes in terms of therapeutic indication. Clinical equivalence among the set of treatments we use in our model was confirmed by two experts (a lung specialist and the head of the pharmacy department of the medical school at our university).

In the coming weeks, while meeting with other asthma patients, Dr. Jones recalls the experiences of patients who tried Symbicort and continuously updates his quality beliefs about the new brand. He recalls, from his medical training, that he should consider the experiences of all his patients with the new drug (as large a sample as possible). Yet Dr. Jones seems to recall the complaints of Mrs. Smith and Mr. Miller much more readily than the feedback provided by other patients. The complaints of Mrs. Smith and Mr. Miller are, therefore, particularly influential in Dr. Jones' quality belief formation about Symbicort and in his adoption decision.

The objective of our model is to extend the Bayesian learning framework to enable it to accommodate the type of salience effects that Dr. Jones experiences while he learns about the quality of Symbicort. We find that a salience effect is indeed present, and it affects physician learning. As to its magnitude, we find that feedback from patients who are salient in the physician's mind receives between 7 and 10 times more weight than feedback from other patients. Physicians' choices also exhibit within-patient persistence, suggesting that physicians (and patients) perceive a cost to switching treatment.

Finally, our model brings valuable insights for firms launching new therapies, a high-research priority area according to life sciences managers and marketing scholars (Stremersch 2008, Stremersch and Van Dyck 2009). In particular, salience slows down the adoption of new treatments. Using counterfactual simulations, we show that AstraZeneca could have increased its market share by as much as 8.5 percentage points by eliminating salience for its new brand (Symbicort). Furthermore, if a policy maker is able to reduce salience across all treatments, physicians adopt newer combination treatments significantly faster. We explore the managerial and policy implications of these findings.

2. Salience in Physician Learning

We now discuss the antecedents and consequences of salience of patients subject to treatment switching. We also discuss other drivers of prescription choices.

2.1. Antecedents of Salience of Patients Subject to Treatment Switching

Salience may result from medical ethics, cognition, and emotion, which we discuss next, each in turn. First, medical practice rests on strong ethical foundations. The motto *Primum non nocere* (first, do no harm) is a central ethical principle that guides the practice of medicine (Brewin 1994). Thus, physicians are under ethical and legal pressure to avoid any (unnecessary) risk that could potentially harm patients. This pressure may enhance the salience of patients who did not

react to a treatment as expected and, consequently, had to be switched to a substitute treatment (an experience the physician wants to minimize in the future).

Second, psychological and neurological research suggests that we react more strongly to undesirable outcomes than to desirable ones (for an overview, see Baumeister et al. 2001 and Rozin and Royzman 2001). During learning and information processing, negative information receives more attention and more elaboration than positive information (Baumeister et al. 2001). Switching a patient to a clinically equivalent treatment is an undesirable outcome for the physician, as it means the patient's reaction to the treatment was different from what the doctor had hoped. Thus, the salience of switching patients may have a cognitive rationale.

Third, treatment switching usually reveals disconfirmation of physician or patient expectations from a treatment. Along the reasoning of Oliver (1993), disconfirmation of expectations provokes not only a cognitive response but also a negative affective response. Moreover, treatment switching can be seen by some patients as a correction to a prior decision and, consequently, perceived by the physician as a threat to her reputation. The ensuing negative emotions alert the physician to the need to eliminate or reduce the trigger of such threats (Taylor 1991). Salience of switching patients will then emerge as a natural consequence of these affective responses and of the human tendency to respond more strongly to negative than to positive emotions (Cacioppo and Gardner 1999).

2.2. Consequences of Salience of Patients Subject to Treatment Switching

Salience interferes with belief formation through the dynamics of over- and underconfidence about different information signals. Griffin and Tversky (1992) show that when weighting evidence, humans tend to overreact to the extremeness and vividness of information (*strength*) irrespective of its predictive validity (*weight*). Compared with a normative statistical model—where evidence and prior beliefs are integrated using Bayes' rule—experimental subjects in their studies were overconfident about evidence when strength was high and weight was low, but underconfident when strength was low and weight was high.

According to Griffin and Tversky (1992), the overconfidence about salient information results from the combination of two cognitive shortcuts: anchoring and adjustment, and representativeness (Tversky and Kahneman 1974). A third cognitive shortcut that can contribute to the influence of salience in belief formation is the availability heuristic (Tversky and Kahneman 1974). Experimental and survey-based research indeed suggests that these heuristics interfere with medical decisions (Klein 2005, Poses and Anthony 1991, Redelmeier 2005).

As a result, we hypothesize that physicians give extra weight to feedback provided by easier-to-recall patients, i.e., those who are switched to an alternative treatment. The influence of feedback provided by switching patients will thus be systematically stronger than what is predicted by a pure Bayesian learning model.

2.3. Other Drivers of Prescription Choices

In addition to quality beliefs, other effects might also drive treatment choices. First, we expect patients to face a switch cost whenever they change treatment, a cost that the physician takes into account in her treatment choices. This switch cost is estimated controlling for quality perceptions; thus it captures a nonquality-based persistence, e.g., the psychological impact of changing treatments (see also Chan et al. 2010). Second, treatments can have serious side effects, so physicians may be risk averse in their treatment choices. Thus, we allow for risk aversion in our model specification. Note that substantial debate exists on physicians' risk attitude, with some studies finding physicians to be risk neutral (Chintagunta et al. 2009, Narayanan et al. 2005, Narayanan and Manchanda 2009), whereas others find physicians to be risk averse (Ching and Ishihara 2010, Coscelli and Shum 2004, Crawford and Shum 2005). Third, we control for marketing effects using a reduced-form approach, i.e., by letting marketing expenditures shift the utility levels of each treatment alternative (for a similar approach, see Chintagunta et al. 2009).

3. Model Specification

In this section, we first lay down the pure Bayesian learning component of our model. Next, we extend this specification by introducing salience in a quasi-Bayesian fashion. This structure clarifies that our quasi-Bayesian model nests its pure Bayesian counterpart. We close the section with the utility specification. Whenever we use mathematical symbols, i indexes physicians ($i = 1, \dots, N$), p indexes patients ($p = 1, \dots, P_i$ being the patients of physician i), k indexes encounters ($k = 1, \dots, K_i$ being the encounters of physician i), and j indexes treatments ($j = 1, \dots, J$).

3.1. Pure Bayesian Learning Framework

We define mean quality of treatment j for physician i (Q_{ij}) as a general attribute that summarizes how well, across all patients of physician i , the treatment provides symptomatic relief (i.e., relief during asthma attacks) and maintains patient health (i.e., avoids recurrence of such attacks) while avoiding severe side effects (for a similar definition, see, e.g., Narayanan et al. 2005). However, a certain treatment j will not work equally well for every patient. Therefore, we

explicitly model patient heterogeneity, i.e., the across-patient variability of treatment quality ($\sigma_{q,ipj}^2$), in line with the work of Chintagunta et al. (2009). Thus, we define the true quality of treatment j for patient p , visiting physician i , as the sum of the true mean quality of treatment j across all patients of physician i and a patient-specific deviation from this mean quality; i.e.,

$$Q_{ipj} = Q_{ij} + q_{ipj}, \quad \text{with } q_{ipj} \sim N(0, \sigma_{q,ipj}^2). \quad (1)$$

Next, we assume that, at the start of our data, each physician has a prior (uncertain) belief about Q_{ij} , treatment j 's mean quality, and about q_{ipj} , the patient-treatment idiosyncratic deviation. We specify a normal distribution for these initial beliefs:

$$Q_{0,ij} \sim N(\bar{Q}_{0,ij}, \sigma_{Q_{0,ij}}^2), \quad (2)$$

$$q_{0,ipj} \sim N(\bar{q}_{0,ipj}, \sigma_{q_{0,ipj}}^2). \quad (3)$$

Here, we assume that $\bar{q}_{0,ipj} = 0$, i.e., that when seeing a new patient, physician i believes that the quality of treatment j , for that particular patient, is equal to the population mean. We also assume rational expectations, a common practice in Bayesian learning models (e.g., Crawford and Shum 2005, Narayanan and Manchanda 2009). Under this assumption physicians have correct initial beliefs about mean quality and quality dispersion across patients, even though they do not know the quality of each treatment for a specific patient. In our model this assumption means that $\bar{Q}_{0,ij} = Q_{ij}$ and $\sigma_{q_{0,ipj}}^2 = \sigma_{q,ipj}^2$. Starting from these prior beliefs, physicians learn about treatment quality in order to (i) reduce the uncertainty surrounding their mean quality belief and (ii) learn about each patient's idiosyncratic deviation from a treatment's mean quality.

We assume physicians learn from their clinical experience, i.e., from the feedback provided by their patients. At the start of each medical encounter, a patient provides a feedback signal about the treatment that was prescribed in her last encounter. These feedback signals are truthful but noisy. That is, if at encounter k physician i receives a feedback signal from patient p about treatment j , we assume this feedback signal to be normally distributed:

$$F_{ipj,k} | Q_{ipj} \sim N(Q_{ipj}, \sigma_{F,i}^2). \quad (4)$$

Note that Bayesian learning guarantees that, even though patients only provide feedback about the last treatment their physician prescribed them, physician i 's treatment choices are influenced by the feedback received from all patients on all treatments. The information set of physician i , at encounter k , is then the clinical history of all her patients, which can be summarized by the average of each patient's feedback

signals up to and including encounter k (denoted as $\bar{F}_{ipj,k}$).

Our assumptions of normally distributed prior beliefs and feedback signals guarantee that physician i 's posterior beliefs are also normally distributed. Specifically, physician i 's posterior belief, at encounter k , about the mean (across-patient) quality of treatment j is

$$\bar{Q}_{ij,k} \sim N(\bar{Q}_{ij,k}, \sigma_{\bar{Q}_{ij,k}}^2). \quad (5)$$

The mean and the variance in Equation (5) result from the assumption that physician i integrates each patient's clinical history ($\bar{F}_{ipj,k}$) with her prior beliefs according to Bayes' rule (Chintagunta et al. 2009; DeGroot 1970; see Online Appendix A for the derivation. An electronic companion to this paper is available as part of the online version that can be found at <http://mktsci.pubs.informs.org/>). That is, denoting the number of feedback signals provided by patient p to physician i about treatment j up to and including encounter k by $n_{ij,k}^p$, we can write

$$\bar{Q}_{ij,k} = \frac{\sigma_{Q_{0,ij}}^2}{\sigma_{Q_{0,ij}}^2 + \sum_p \frac{n_{ij,k}^p \cdot \sigma_{Q_{0,ij}}^2}{\sigma_{F,i}^2 + n_{ij,k}^p \cdot \sigma_{q,ipj}^2}} \cdot \bar{Q}_{0,ij} + \sum_p \frac{n_{ij,k}^p \cdot \sigma_{Q_{0,ij}}^2}{\sigma_{F,i}^2 + n_{ij,k}^p \cdot \sigma_{q,ipj}^2} \cdot \bar{F}_{ipj,k}, \quad (6)$$

where

$$\sigma_{\bar{Q}_{ij,k}}^2 = \left[1/\sigma_{Q_{0,ij}}^2 + \sum_p \frac{n_{ij,k}^p}{(\sigma_{F,i}^2 + n_{ij,k}^p \cdot \sigma_{q,ipj}^2)} \right]^{-1}. \quad (7)$$

Similarly, physician i 's posterior belief, at encounter k , about patient p 's idiosyncratic deviation from this mean quality is defined as

$$\bar{q}_{ipj,k} \sim N(\bar{q}_{ipj,k}, \sigma_{\bar{q}_{ipj,k}}^2). \quad (8)$$

The mean, in Equation (8), results from physician i 's Bayesian updating of her initial prior belief about this deviation ($\bar{q}_{ipj,0}$) with the observed difference between the mean of patient p 's feedback signals about treatment j up to and including encounter k ($\bar{F}_{ipj,k}$) and physician i 's belief, at encounter k , about the mean quality of treatment j across patients ($\bar{Q}_{ij,k}$); i.e.,

$$\bar{q}_{ipj,k} = \frac{\sigma_{q,ipj,k}^2}{\sigma_{q_{0,ipj}}^2} \cdot \bar{q}_{0,ipj} + \frac{n_{ij,k}^p \cdot \sigma_{q,ipj,k}^2}{\sigma_{F,i}^2} \cdot (\bar{F}_{ipj,k} - \bar{Q}_{ij,k}), \quad (9)$$

where

$$\sigma_{\bar{q}_{ipj,k}}^2 = (1/\sigma_{q_{0,ipj}}^2 + n_{ij,k}^p/\sigma_{F,i}^2)^{-1}. \quad (10)$$

The expected quality (i.e., the mean belief of physician i) of prescribing treatment j to the patient visiting at occasion k is obtained by adding Equations (6) and (9). We now introduce salience and, afterward, discuss dynamics in physicians' uncertainty about the quality of each treatment.

3.2. Introducing Salience

We modify Bayesian learning in order to incorporate the different roles of information weight and information salience, or strength as referred by Griffin and Tversky (1992). Specifically, *weight* depends on (i) the variance of patient feedback signals ($\sigma_{F,i}^2$), (ii) the variances of physician i 's prior quality beliefs ($\sigma_{Q0,ij}^2$ and $\sigma_{q0,ipj}^2$), and (iii) the number of feedback signals provided by each patient. We introduce a salience parameter, $\omega_{ipj,k}$, which quantifies the *extra* weight—when compared with the pure Bayesian learning weight—given by physician i to the feedback of a patient, p , who was subject to treatment switching. We define $SWITCH_{ipj,k}$ as a dummy variable that assumes the value one if, before encounter k , patient p has been switched away from treatment j and has not yet been prescribed treatment j again (until the start of encounter k). We then introduce the impact of salience in physician i 's posterior belief about the mean quality of treatment j as follows:

$$\begin{aligned} \bar{Q}_{ij,k}^\omega &= \frac{\psi_{Q,\omega,ij,k}^2}{\sigma_{Q0,ij}^2} \cdot \bar{Q}_{0,ij} \\ &+ \sum_p \frac{\psi_{Q,\omega,ij,k}^2 \cdot n_{ij,k}^p \cdot (1 + \omega_{ipj,k} \cdot SWITCH_{ipj,k})}{\sigma_{F,i}^2 + n_{ij,k}^p \cdot \sigma_{q,ipj}^2} \\ &\cdot \bar{F}_{ipj,k}, \end{aligned} \quad (11)$$

where

$$\begin{aligned} \psi_{Q,\omega,ij,k}^2 &= \left[1/\sigma_{Q0,ij}^2 + \sum_p n_{ij,k}^p \cdot (1 + \omega_{ipj,k} \cdot SWITCH_{ipj,k}) \right. \\ &\left. / (\sigma_{F,i}^2 + n_{ij,k}^p \cdot \sigma_{q,ipj}^2) \right]^{-1}. \end{aligned} \quad (12)$$

Please note that the expression in Equation (12) is not the variance of physician i 's belief about the mean quality of j as would be the case in a pure Bayesian setting. This is because in our model, physicians' beliefs are affected by salience, and hence, they learn in a quasi-Bayesian manner (see Boulding et al. 1999, Rabin and Schrag 1999). After a treatment switch, a physician changes the relative weight she gives to the feedback of the switching patient about the abandoned treatment vis-à-vis her prior belief and the feedback of other patients. The resulting posterior belief will, therefore, necessarily depart from a pure Bayesian belief. Given that Bayes' rule is the optimal way to learn, we expect this departure to be manifested in slower learning. In sum, we specify a quasi-Bayesian learning model that incorporates salience effects but is equivalent to a pure Bayesian learning model in case $\omega_{ipj,k} = 0$.

Note that salience predicts a systematic misinterpretation of patient feedback about the quality

of abandoned treatments. In contrast, salience does not predict any systematic bias in physicians' belief updating for any other treatment alternative. Therefore, the physician's choice of a new treatment will always reveal that physician's preference for that treatment, irrespective of whether the patient was switched from another treatment or is a new patient.³

To test whether salience is a temporary or permanent phenomenon, and whether its magnitude changes with temporal distance from the focal switch, we operationalize salience using three parameters:

$$\omega_{ipj,k} = \omega_{0,i} \cdot \lambda_{ipj,k}^\omega + \omega_{\infty,i} \cdot (1 - \lambda_{ipj,k}^\omega), \quad (13)$$

with the weight given to the immediate magnitude of salience ($\omega_{0,i}$), vis-à-vis the long-term magnitude of salience ($\omega_{\infty,i}$), decreasing over time as follows: $\lambda_{ipj,k}^\omega = 2/\{1 + \exp[\lambda_i^\omega \cdot (\tau(k) - \tau(t_{ipj}^{sw}))]\}$. Here, $\tau(k)$ denotes the calendar date of encounter k , and $\tau(t_{ipj}^{sw})$ denotes the calendar date of the last occasion when patient p had to be switched away from treatment j in favor of one of the clinically equivalent treatments. The rate of decay is governed by the parameter λ_i^ω , which is assumed to be positive.

If salience interferes with the formation of mean quality beliefs, as specified in Equation (11), it will also interfere with physicians' beliefs about patient-treatment idiosyncratic deviations; i.e.,

$$\tilde{q}_{ipj,k}^\omega = \frac{\sigma_{q,ipj,k}^2}{\sigma_{q0,ipj}^2} \cdot \tilde{q}_{0,ipj} + \frac{n_{ij,k}^p \cdot \sigma_{q,ipj,k}^2}{\sigma_{F,i}^2} \cdot (\bar{F}_{ipj,k} - \bar{Q}_{ij,k}^\omega). \quad (14)$$

The posterior belief of a quasi-Bayesian physician about the quality of treatment j for patient p visiting at encounter k is then again the sum of her posterior belief about the mean quality of treatment j ($\bar{Q}_{ij,k}^\omega$) and her posterior belief about patient p 's idiosyncratic deviation from this mean ($\tilde{q}_{ipj,k}^\omega$); i.e.,

$$\tilde{Q}_{ipj,k}^\omega \sim N(\bar{Q}_{ij,k}^\omega + \tilde{q}_{ipj,k}^\omega, \sigma_{Q,\omega,ipj,k}^2). \quad (15)$$

The posterior variance in Equation (15) describes how the uncertainty about the quality of treatment j for patient p , for a quasi-Bayesian physician, evolves over time; i.e.,

$$\begin{aligned} \sigma_{Q,\omega,ipj,k}^2 &= \text{var}(\tilde{Q}_{ipj,k}^\omega) \\ &= \text{var}(\bar{Q}_{ij,k}^\omega) + \text{var}(\tilde{q}_{ipj,k}^\omega) \\ &\quad + 2 \cdot \text{cov}(\bar{Q}_{ij,k}^\omega, \tilde{q}_{ipj,k}^\omega). \end{aligned} \quad (16)$$

We provide the derivation of $\sigma_{Q,\omega,ipj,k}^2$ in Online Appendix B of the electronic companion. If the estimated salience ($\omega_{ipj,k}$) is different from zero, it provides evidence in favor of our hypothesized deviation

³ We thank the associate editor and an anonymous reviewer for pointing out the need to clarify this issue.

from Bayesian updating. As a final note, the distribution of physician i 's beliefs, across all patients at encounter k , results in a structure like the one derived in Chintagunta et al. (2009).

3.3. Utility Specification

In line with previous research (e.g., Erdem and Keane 1996, Narayanan and Manchanda 2009), we assume that, at each encounter k , physician i chooses the treatment j that, according to her beliefs, maximizes the expected utility of patient p , which is given by

$$U_{ipj,k} = \bar{Q}_{ij,k}^{\omega} + \bar{q}_{ipj,k}^{\omega} - \frac{1}{2} \cdot r_i \cdot \sigma_{Q,\omega,ipj,k}^2 + \delta_i \cdot \text{LASTCHOICE}_{ipj,k} + \text{MARKETING}_{ij,k} + \varepsilon_{ipj,k}. \quad (17)$$

In this specification r_i is the absolute risk-aversion coefficient, which measures each physician's risk attitude. A positive r_i indicates that physicians are risk averse, i.e., less inclined to prescribe a treatment when quality uncertainty is larger. Quality uncertainty enters the utility function via $\sigma_{Q,\omega,ipj,k}^2$, the posterior variance of $\bar{Q}_{ipj,k}^{\omega}$, as defined in Equation (16). The dummy variable $\text{LASTCHOICE}_{ipj,k}$ assumes the value one if the physician prescribed treatment j to patient p in their last encounter and δ_i is a parameter capturing switch costs, i.e., a propensity of physician i to prescribe to patient p the same treatment j that had been prescribed in their last encounter. To control for the impact of marketing efforts, we use $\text{MARKETING}_{ij,k}$, which is a flexible function of the market-level marketing expenditures.⁴ Finally, $\varepsilon_{ipj,k}$ is an error term capturing unobserved drivers of utility at encounter k . We assume these errors to be normally distributed and allow for between-treatment covariation; i.e., $\varepsilon_{ip,k[1:x]} \sim N(\mathbf{0}, \Sigma)$. Our data provide a natural structure for the correlations across treatments.⁵

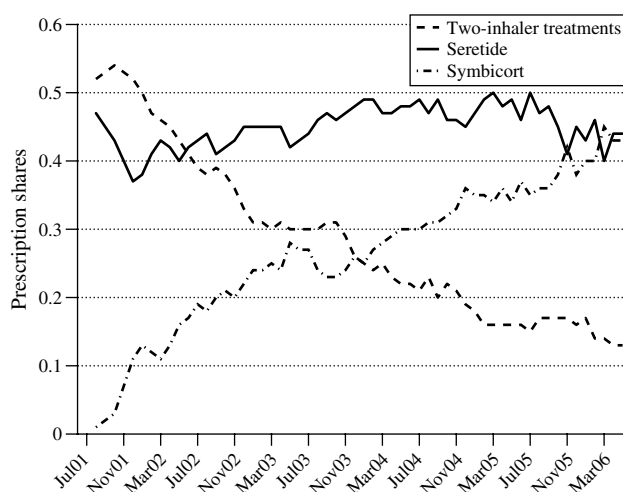
4. Data

The market we study is the obstructive airways diseases category (i.e., asthma and chronic obstructive pulmonary disease, or COPD)—in particular, the class of inhaled corticosteroids (ICSs) plus long-acting β_2 -agonist (LABA) combinations, in The Netherlands.

⁴ In short, in $\text{MARKETING}_{ij,k}$ we integrate two components: (i) temporal responsiveness to marketing actions, assumed equal across physicians, and molecules but changing over time; and (ii) physician-specific marketing responsiveness, which is assumed constant over time. We will discuss in greater detail our specification of marketing after we have introduced our data.

⁵ We distinguish between (i) treatment- and encounter-specific shocks, which are independent across treatments; and (ii) ingredient- or administration-specific shocks, which can affect all treatments that share a certain molecular ingredient or route of administration.

Figure 1 Prescription Shares (Three-Month Moving Average)



By 2017, global sales of medications for asthma and COPD are expected to reach \$25 billion with ICS-plus-LABA combinations becoming the leading class in value (Datamonitor 2008). ICS-plus-LABA combinations are recommended for patients with moderately severe asthma (Global Initiative for Asthma (GINA) 2008) and chronic obstructive pulmonary disease (Calverley et al. 2007).

We model treatment choice among eight clinically equivalent treatments. These include six two-inhaler combinations of the three ICSs (beclomethasone, budesonide, and fluticasone) and two LABAs (formoterol and salmeterol) recommended by clinical guidelines (GINA 2008); and two newer single-inhaler brands—GlaxoSmithKline's Seretide (fluticasone + salmeterol; approved in 1999; branded as Advair in the United States) and AstraZeneca's Symbicort (budesonide + formoterol; approved in 2001). Our data contain the introduction of Symbicort and cover a period of growing popularity, among physicians, of ICS-plus-LABA combinations, which is an ideal setting to model physician learning about the quality of different treatment alternatives within this category.

Figure 1 depicts the evolution of prescription shares over time, in our sample, of the older two-inhaler treatments and the two newer combination brands. Roughly one year after the start of the observation period, Seretide had a higher prescription share than all the two-inhaler treatments together. Symbicort eventually reached a prescription share similar to Seretide, but only five years after its entry.

We obtained electronic patient records from July 2001 to June 2006 from the IPCI (Integrated Primary Care Information) database,⁶ a panel of general

⁶ In fact, we have access to prescription data before July 2001, but we were only able to gather marketing data, which we will describe

practitioners, maintained by the school of medicine at our university (for a detailed description, see Vlug et al. 1999). These physicians use *paperless offices*, meaning that the system records the full prescription history of each patient, including all refills. The data from this panel are often used for research publications in medicine and pharmaco-epidemiology. Usage of the data is supervised by a board of medical professionals, and linking the data to other sources at the individual physician level is prohibited.

The panel contains both single- and multiphysician practices. To ensure that we model belief formation using all the relevant clinical experience for each physician, we only use data on single-physician practices. The data contain 2,398 patients across 22 physicians and 12,186 prescription choices (of ICS-plus-LABA treatments).⁷ We obtained data on monthly expenditures on marketing (including detailing, journal advertising, and conferences), for the respective treatments and time period from IMS Health. We use these data to construct, for each treatment and occasion, the marketing variable introduced in Equation (17) as follows:

$$\begin{aligned}
 & \text{MARKETING}_{ij,k} \\
 &= \sum_{l=0}^L \{ \beta_{1,l}^{\text{Mkt}} \cdot [I(\text{two-inhalers}) \cdot (\beta_{2,i}^{\text{ICS}} \cdot \ln(\text{MKT}_{j,m(k)-l}^{\text{ICS}} + 1) \\
 & \quad + \beta_{2,i}^{\text{LABA}} \cdot \ln(\text{MKT}_{j,m(k)-l}^{\text{LABA}} + 1)) \\
 & \quad + (1 - I(\text{two-inhalers})) \cdot \beta_{2,i}^{\text{Comb.}} \\
 & \quad \cdot \ln(\text{MKT}_{j,m(k)-l}^{\text{Comb.}} + 1)] \}. \tag{18}
 \end{aligned}$$

Here, $L = 6$ represents the number of lagged monthly marketing expenditures that, in our model, affect a treatment's utility; $I(\text{two-inhalers})$ is an indicator function assuming the value one if treatment j combines the preventive ICS and reliever LABA molecules in two distinct inhalers, and value zero if these two molecules are combined in the same inhaler; and $m(k)$ indicates the calendar month of encounter k (contemporaneous marketing expenditures, i.e., when $l = 0$, are adjusted for the timing of the encounter within a given month).

Table 1 presents a switching matrix among two-inhaler treatments, Seretide and Symbicort. It shows that physicians tend to switch patients away from

Table 1 Switching Matrix

From	To			Switch rate (%)	Total number of prescriptions
	Two-inhaler treatments	Seretide	Symbicort		
Two-inhaler treatments	67	140	84	6.24	3,592
Seretide	43	0	50	1.69	5,506
Symbicort	35	39	0	2.40	3,088

Table 2 Descriptive Statistics—Patient Visits

Measure	Mean	SD	Min	Max
No. of prescription occasions per patient	5.08	5.85	1	50
No. of spells per patient	1.22	0.68	1	10
Spell length	4.16	4.84	1	41

two-inhaler treatments to Seretide (140 switches; i.e., 30.1% of the 458 switches) or to Symbicort (84 switches; i.e., 18.3% of the 458 switches). Yet we also observe 43 switches from Seretide and 35 from Symbicort to two-inhaler treatments (i.e., 78 switches in total, or 17% of the 458 switches) as well as between Seretide and Symbicort. Column “Switch rate (%)” shows the switching rate as a percentage of the total number of prescriptions (which are, in turn, displayed under column “Total number of prescriptions”).

Let us now define a spell as a sequence of consecutive prescriptions of the *same treatment* (Crawford and Shum 2005). Table 2 shows descriptive statistics for our data. On average, a patient receives five prescriptions. The average number of spells per patient is 1.22 (we observe a total of 2,926 spells across the 2,398 patients), and, on average, each spell consists of 4.16 prescriptions. The average number of spells is very close to previous studies (Chintagunta et al. 2009, Crawford and Shum 2005). The mean length of each spell is larger than prior studies because we deal with patients having a moderate to severe chronic disease.

5. Estimation and Identification

5.1. Estimation

We estimate our model in a Bayesian fashion using a Markov chain Monte Carlo (MCMC) approach. We sample the parameters from their posterior distributions using a Gibbs sampler (see Casella and George 1992 for a review) together with data augmentation that allows us to sample the latent utilities and patient feedback alongside the model parameters (Tanner and Wong 1987). In addition, in line with Narayanan and Manchanda (2009), we use a hierarchical Bayes' structure to model unobserved physician heterogeneity. We adapt McCulloch and Rossi's (1994) Gibbs sampler for hierarchical multinomial probit models to

shortly, from July 2001 onward. Still, we used data before July 2001 to initiate the switch cost variable. Moreover, although the formal approval of Symbicort was in January 2001, in our data, only two prescriptions are recorded before July 2001.

⁷ To avoid concerns with patient dropout in our data, we have compared the prescription shares in the full sample with the shares among patients who have dropped off the panel and found no significant differences.

account for Bayesian or quasi-Bayesian learning. The main difference, besides having to use data augmentation to sample patient feedback is that we do not have a closed-form solution for the posterior distributions of (i) the variances characterizing physicians' initial uncertainty and patient heterogeneity ($\{\sigma_{Q0,ij}^2\}$ and $\{\sigma_{q,ipj}^2\}$), (ii) the variance of patients' feedback ($\sigma_{F,i}^2$), and (iii) the saliency parameters ($\omega_{0,i}$, $\omega_{\infty,i}$, and λ_i^ω). To sample these parameters, we apply a Metropolis-Hastings step (Chib and Greenberg 1995) within our Gibbs sampler. We specify proper but diffuse priors for all parameters. The exact implementation of our Gibbs sampler is given in Online Appendix C of the electronic companion. We let all chains converge and use 5,000 subsequent draws to obtain parameter estimates.

5.2. Identification

The structure of Bayesian learning and the dynamics in prescription shares—including the introduction of a new treatment (Symbicort)—help us in identification of the learning parameters. When Symbicort is introduced, physicians are uncertain about its quality, and learning helps them reduce such uncertainty over time. The velocity of this reduction depends on the noise in feedback signals ($\sigma_{F,i}^2$) and on the variances characterizing prior quality uncertainty and patient heterogeneity ($\sigma_{Q0,ij}^2$ and $\sigma_{q,ipj}^2$).

To identify $\sigma_{Q0,ij}^2$ and $\sigma_{q,ipj}^2$, we rely on the attractiveness of a treatment for new versus old patients. Bayesian updating guarantees that the uncertainty surrounding the mean quality of a treatment ($\sigma_{Q0,ij}^2$) tends to zero after a large enough number of signals. At this point, the reluctance of a physician to prescribe that treatment to a new patient (which also does not depend on switch costs) enables identification of $\sigma_{q,ipj}^2$.

The assumption that physicians have rational expectations enables the dynamics in the choices of treatments with higher versus lower quality uncertainty to identify risk aversion (r_i) and switch costs (δ_i). If quality expectations are on average correct, relative sluggishness in prescribing treatments with higher associated uncertainty to new patients is driven by risk aversion ($r_i > 0$). Sluggishness in switching revisiting patients to treatments that the physician has already adopted for new patients enables identification of the switch cost parameter (δ_i). Thus, an overall unwillingness to try more uncertain treatments identifies risk aversion, whereas an unwillingness to switch revisiting patients away from a certain treatment identifies switch costs.

The saliency parameters ($\omega_{0,i}$, $\omega_{\infty,i}$, and λ_i^ω) are identified by systematic changes in behavior triggered by the decision to switch a patient to a clinically equivalent alternative. For instance, if a physician

starts adopting Symbicort to several patients at a certain pace but, after switching a patient away from Symbicort, slows down this adoption process more than what Bayes' rule would predict, this reduction in the speed of adoption is captured by $\omega_{0,i}$. If the strength of this effect changes over time, our model will capture such dynamics through λ_i^ω and $\omega_{\infty,i}$. Please note that it is not possible to identify the sign of the decay parameter separately from the levels of the two saliency parameters. We avoid this identification issue by restricting λ_i^ω to be positive.

For the marketing parameters (β_1^{Mkt} and $\beta_{2,i}^{Mkt}$), identification is straightforward. Controlling for learning and switch costs, the effect of marketing efforts on the attractiveness of alternative treatments is identified by the responsiveness, in terms of prescription choices, of physicians to variations in the marketing effort variables. For identification purposes, we assume that the temporal marketing responsiveness parameters add up to one.

Unrestricted multinomial probit models suffer from additional identification issues because choice probabilities are invariant to location or scale transformations of the latent utilities (Rossi et al. 2005). Hence, we normalize the scale and location of the utility levels by restricting the quality of a reference alternative—the two-inhaler combination of fluticasone and salmeterol—to zero and the variance of the error term of the reference alternative to one. Note that the latter especially has implications for comparability of estimation results across models (see Swait and Louviere 1993). Estimates of the utility levels, variances, marketing effects, as well as risk aversion and switch costs, will be affected by the restriction in the variance of the error terms and by the amount of unexplained variation actually present in the behavior under consideration.

As a final note, this type of model is demanding in terms of identification. To guarantee that our results are robust, we have run our focal models using a different set of priors. Even though we made priors much more diffuse, by increasing prior variances by a factor of 10, results were largely unchanged, which suggests that identification of our model is achieved without relying on information contained in the priors. Furthermore, we have also simulated data according to our model and were able to recover the parameters very well.

6. Results

Posterior estimates of the relevant parameters are obtained directly from the sample of MCMC draws. To isolate the contribution of saliency, we compare the following models: (M0) a pure Bayesian learning model, (M1) a quasi-Bayesian learning model with

static salience (i.e., $\omega_i = \omega_{\infty, i} = \omega_{0, i}$), and (M2) a quasi-Bayesian learning model with dynamic salience. Following the suggestion of Rossi et al. (2005) to focus on the log-likelihood to verify convergence, we apply Raftery and Lewis's (1992) *I-stat* and Geweke's (1992) convergence tests on the log-likelihood for all three models, which confirmed that the chains have converged.

Next, we compare the fit of the models using log-marginal densities (LMDs) and log-Bayes' factors (Kass and Raftery 1995). The two quasi-Bayesian learning models ($LMD_{M1} = -19,518$ and $LMD_{M2} = -19,509$) clearly outperformed the pure Bayesian learning benchmark ($LMD_{M0} = -27,845$). This provides strong evidence that any of the quasi-Bayesian learning models is a posteriori more likely than the pure Bayesian learning model, assuming equal prior probabilities for all models. The log-Bayes' factor of the model with dynamic salience (M2) with respect to the model with nondynamic salience (M1) is also above five, the threshold suggested by Kass and Raftery (1995) for strong evidence in favor of the best-fitting model, which supports dynamics in salience effects. Finally, including two different patient feedback signal variances (one for the first encounter and another for subsequent encounters) to accommodate experience effects in the patient-physician relationship did not improve the quasi-Bayesian learning models (M1 and M2) based on log-Bayes' factors. We now turn to the parameter estimates and their interpretation.

6.1. Parameter Estimates

6.1.1. Salience. A key finding from our model is the strong salience effect triggered by the decision to switch a patient to an alternative treatment option (see first three rows of Table 3). When learning about the quality of a treatment, feedback from patients subject to treatment switching receives between 7 and 10 times more weight ($\bar{\omega}_{0, i} = 9.05$, $SD = 0.432$; $\bar{\omega}_{\infty, i} = 6.31$, $SD = 0.435$; and $\bar{\lambda}_i^{\omega} = 1.27$, $SD = 0.566$)⁸ than a pure Bayesian learning model would predict.

We now turn to the dynamics in salience effects. We fix $\bar{\lambda}_i^{\omega}$ at its mean and compute the magnitude of salience since the time of a switch until one year after the switch based on the medians of $\bar{\omega}_{0, i}$ and $\bar{\omega}_{\infty, i}$. We find that 56% of the total decay from the immediate level of salience ($\bar{\omega}_{0, i}$) to its steady-state level ($\bar{\omega}_{\infty, i}$) occurs in one year's time. Thus, we find evidence for a significant but slow decay in salience.

⁸ The values of 7 and 10 are obtained by adding one to our estimates of $\bar{\omega}_{\infty, i}$ and $\bar{\omega}_{0, i}$, as the impact of salience in the utility is determined by $(1 + \omega_{ipj, k} \cdot SWITCH_{ipj, k})$. We use SD to denote the standard deviation across all the MCMC draws used for posterior inference.

Table 3 Parameter Estimates (Salience, Switch Costs, Risk Aversion, and Feedback Error)

Parameter	Posterior median [95% credible intervals]	Across-physician std. dev.	Within-physician std. dev.
Immediate salience effect ($\bar{\omega}_{0, i}$)	9.05 [8.18; 9.89]	1.21	1.04
Long-run salience effect ($\bar{\omega}_{\infty, i}$)	6.31 [5.43; 7.11]	1.56	1.05
Salience decay ($\bar{\lambda}_i^{\omega}$)	1.27 [0.14; 2.28]	2.18	1.32
Switch costs ($\bar{\delta}_i$)	2.78 [2.57; 3.03]	0.18	0.16
Absolute risk aversion (\bar{r}_i)	0.60 [-0.53; 1.76]	0.86	0.60
Patient feedback error ($\bar{\sigma}_{F, i}^2$)	0.66 [0.53; 0.81]	0.39	0.14

Notes. The estimates reported in the second column are the medians, across all MCMC draws, of the population mean parameter in the second level of our hierarchical model (i.e., the mean in the random coefficients distribution). In parentheses, we report the 2.5th and the 97.5th percentiles of the distribution of these MCMC draws. In the last two columns we report the across-physician standard deviations (the standard deviation of the physician-specific means of each parameter) and within-physician standard deviations (the mean of the physician-specific standard deviations of each parameter), in line with Narayanan and Manchanda (2009).

We also find significant physician heterogeneity in salience effects. To determine whether physician heterogeneity is significant, we follow Narayanan and Manchanda's (2009) approach of contrasting each parameter's across-physician standard deviations with the within-physician standard deviations. If the physician-specific 95% credible intervals for a certain parameter do not overlap, then the across-physician standard deviation of that parameter needs to be larger than the corresponding within-physician standard deviation. From the last two columns of Table 3, we can see that the across-physician variation is substantially larger than the within-physician variation, suggesting significant heterogeneity in salience effects.

Salience has two major effects on prescription behavior. First, because it represents a departure from optimal Bayesian learning, it tends to slow down physician learning about the quality of new treatments, which delays its adoption by physicians. Second, in the long run, it benefits treatments that generate fewer switches (i.e., that have higher quality or lower treatment heterogeneity, or that are targeted to patients that will benefit the most from them). In the next section, we will quantify the overall impact of the salience effect on the market.

6.1.2. Switch Costs. The parameter measuring patient switch costs ($\bar{\delta}_i = 2.78$, $SD = 0.121$) suggests

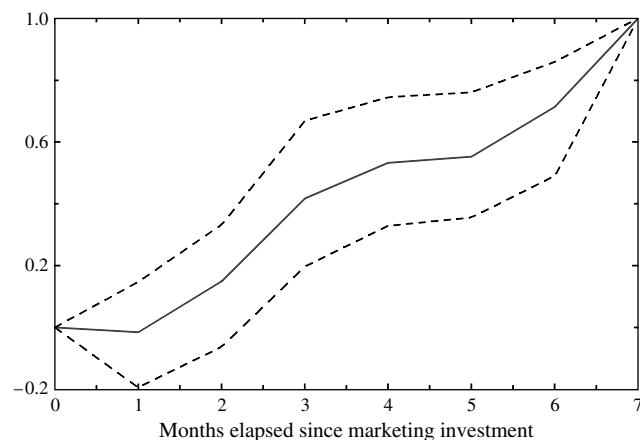
that, on top of uncertainty-driven persistence, physicians exhibit a strong tendency to prescribe the same treatment for a certain patient even when they believe that an alternative treatment could perform better for this specific patient. This finding is in line with the findings of Chan et al. (2010) and Coscelli (2000). Physician heterogeneity in these switch costs seems only marginally significant, as the across-physician and the within-physician standard deviations are close to each other.

6.1.3. Absolute Risk Aversion. The mean risk-aversion parameter is positive, and the standard deviation of the draws is of similar magnitude ($\bar{r}_i = 0.60$, $SD = 0.584$). We computed the percentage of draws indicating risk aversion (i.e., $r_i > 0$) for each of the physicians in our sample and found that all except one physician have the majority of the MCMC draws with positive risk aversion, and for more than half of the physicians, at least 90% of the draws indicate risk aversion. Finally, we find that physicians show significant heterogeneity in their risk attitudes, with the across-physician standard deviation being much larger than the within-physician standard deviation (0.86 versus 0.60), a finding in line with evidence from medicine (Fiscella et al. 2000).

6.1.4. Patient Feedback Error. To understand the magnitude of patients' feedback errors ($\bar{\sigma}_{F,i}^2 = 0.66$, $SD = 0.07$) we simulated physician uncertainty about the mean quality of Symbicort and analyzed how long it takes a physician to reduce such uncertainty. On average, a pure Bayesian physician needs to receive 26 patient feedback signals (each patient providing a single feedback) to reduce her uncertainty by 90%. If the same physician learns in a quasi-Bayesian fashion, i.e., giving more weight to the feedback of salient patients, then she needs 38 signals, all from salient patients, to obtain the same reduction in uncertainty. Thus, as we would expect from the fact that salience represents a deviation from optimal Bayesian learning, salience reduces physicians' speed of learning, which, everything else constant, results in slower adoption. We explore managerial and patient welfare implications of salience in the next two sections.

6.1.5. Marketing Efforts. We now discuss the impact of marketing efforts on treatment utility and choice. We divide marketing responsiveness in two effects: (i) temporal marketing responsiveness, which we use to describe how the effect of pharmaceutical companies' marketing efforts builds up over time (an effect we assume common across physicians and treatments); and (ii) molecule- and physician-specific marketing responsiveness, which is assumed constant over time.

Figure 2 Cumulative Temporal Marketing Responsiveness



Notes. In the horizontal axis we depict the number of months elapsed since the focal marketing investment. For illustrative purposes we start the graph from zero; hence “1” in the horizontal axis refers to the contemporaneous marketing and “7” refers to the sixth lag of temporal marketing responsiveness. In the vertical axis, we depict the sum of the temporal marketing responsiveness parameters up to and including the lag indicated in the horizontal axis. The lines depict the 2.5th (lower dashed line), the median (solid line), and the 97.5th (upper dashed line) percentiles, across all MCMC draws, of $CTMR_L$ (for $L = 1, \dots, 7$).

To describe temporal marketing responsiveness, in Figure 2 we depict the cumulative temporal marketing responsiveness effect since the period a marketing investment is effected until L months have passed since such investment (i.e., $CTMR_L = \sum_{i=0}^{L-1} \beta_{1,i}^{Mkt}$ with $L = 1, \dots, 7$). For identification, the sum of the temporal marketing responsiveness parameters is restricted to one. Hence, the curves in Figure 2 represent the fraction of the total marketing responsiveness that has already affected physician i 's prescription behavior when L months have passed since marketing was expended. We can conclude that marketing effects gradually build up from the first to the seventh month after the investment is made.

In terms of the molecule- and physician-specific marketing responsiveness, our estimates show that marketing efforts to promote Seretide and Symbicort ($\bar{\beta}_{2,i}^{Comb.} = 0.12$, $SD = 0.05$, and 95% credible interval = $[0.03; 0.22]$) significantly drive prescription choices. In contrast, marketing expenditures for ICSs ($\bar{\beta}_{2,i}^{ICS} = -0.01$, $SD = 0.05$, and 95% credible interval = $[-0.11; 0.08]$) and for LABAs ($\bar{\beta}_{2,i}^{LABA} = 0.06$, $SD = 0.05$, and 95% credible interval = $[-0.03; 0.16]$) do not significantly affect prescription behavior, in line with prior research showing the effectiveness of pharmaceutical marketing to be higher for newer than for more mature treatments (Narayanan et al. 2005).

6.1.6. Treatment Characteristics. Table 4, summarizes the posterior medians (and 95% credible intervals) for each treatment's true mean quality (\bar{Q}_{ij}), initial physician uncertainty about the mean quality belief ($\bar{\sigma}_{Q_0,ij}^2$), and patient heterogeneity ($\bar{\sigma}_{q,ipj}^2$).

Table 4 Parameter Estimates: Treatment Quality Perceptions

Treatment alternative	\bar{Q}_{ij}	$\bar{\sigma}_{Q0,ij}^2$	$\bar{\sigma}_{q,ipj}^2$
1—Fluticasone + salmeterol		0.87 [0.68; 1.09]	0.80 [0.60; 0.97]
2—Fluticasone + formoterol	-0.06 [-0.30; 0.22]	0.69 [0.57; 0.85]	1.34 [1.10; 1.67]
3—Beclomethasone + salmeterol	-0.04 [-0.30; 0.17]	1.06 [0.81; 1.41]	0.76 [0.61; 0.90]
4—Beclomethasone + formoterol	-0.24 [-0.46; 0.00]	0.68 [0.56; 0.88]	1.09 [0.91; 1.29]
5—Budesonide + salmeterol	-0.02 [-0.24; 0.18]	0.92 [0.79; 1.07]	0.98 [0.81; 1.16]
6—Budesonide + formoterol	0.14 [-0.10; 0.38]	0.67 [0.57; 0.80]	1.04 [0.88; 1.23]
7—Seretide	0.25 [0.04; 0.46]	0.74 [0.60; 0.94]	0.80 [0.65; 0.93]
8—Symbicort	0.10 [-0.12; 0.33]	1.30 [1.05; 1.61]	0.89 [0.74; 1.03]

Note. Fluticasone + salmeterol is the reference treatment alternative; Seretide contains fluticasone and salmeterol, and Symbicort contains budesonide and formoterol.

In terms of true mean qualities, we find that the fourth treatment alternative (two-inhalers combining beclomethasone and formoterol) is the one with lowest quality, whereas Seretide is perceived as the best treatment, on average. This finding is consistent with the results from a pure Bayesian learning model. In fact, the only relevant difference between the two models is that, in the quasi-Bayesian learning model, Seretide’s quality is significantly higher than the remaining alternatives. In contrast, in the pure Bayesian learning model, the estimate for the mean quality of Seretide was very close to zero.

In terms of face validity, the results from our quasi-Bayesian learning model (M2) are consistent with medical studies, which show that the different treatment alternatives in this category are equivalent in terms of efficacy and side effects (Marks and Ind 2005). The fact that Seretide seems to be perceived by the physicians in our sample as having higher mean quality than the remaining treatments is also consistent with evidence from the industry indicating that AstraZeneca’s initial differentiation strategy—which was to allow patients to adjust the dosing of the ICS’s component—may have been received with skepticism by many physicians, who believed that a fixed dosing of ICS was actually one of the advantages of combination treatments.⁹ These differences indicate that treatments are also characterized by other dimensions, such as dosage, administration method, and convenience (Venkataraman and Stremersch 2007).

⁹ See, for example, Datamonitor’s report named “Symbicort and Seretide’s battle for the respiratory market”: http://www.datamonitor.com/store/News/symbicort_and_seretides_battle_for_the_respiratory_market?productid=489DA887-A5B9-4660-B285-29D22EC64F6A, accessed April 2010.

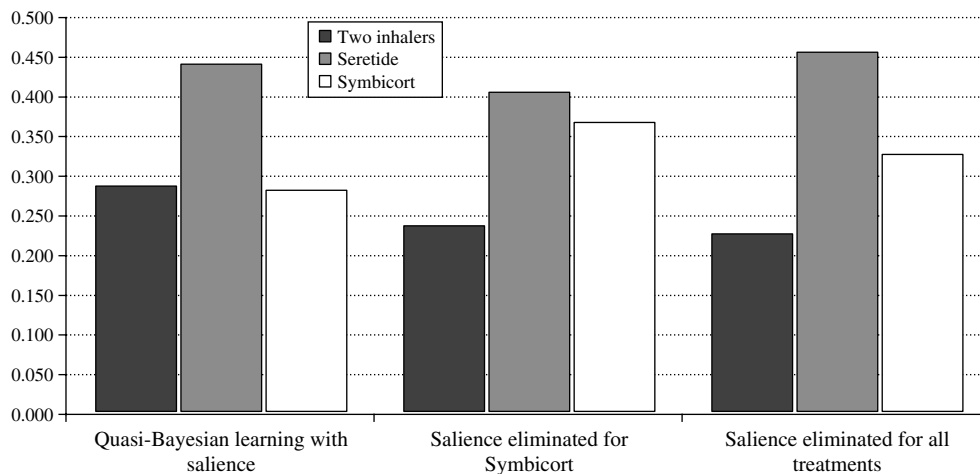
The estimates for initial uncertainty about the mean quality of each treatment also have high face validity in our model. Symbicort, the newest treatment, shows the highest mean quality uncertainty ($\bar{\sigma}_{Q0,iSYMBI}^2 = 1.30$, $SD = 0.142$), whereas Seretide, which had been introduced two years before the start of our data and was, at the time, already the most prescribed treatment, shows significantly lower prior mean quality uncertainty ($\bar{\sigma}_{Q0,iSERE}^2 = 0.74$, $SD = 0.089$, with all draws having $\bar{\sigma}_{Q0,iSYMBI}^2 > \bar{\sigma}_{Q0,iSERE}^2$). Finally, physicians perceive patient heterogeneity ($\bar{\sigma}_{q,ipj}^2$) as similar across all treatments.

7. Effects of Salience on Market Shares

Having established the presence of strong salience effects in physician learning, we now quantify the consequences of this behavioral regularity at the market level. We use the posterior draws from the quasi-Bayesian learning model with dynamic salience (M2) to simulate market shares under two counterfactual experiments: (i) our model with salience set to zero only for Symbicort and (ii) our model with salience set to zero for all treatments. The first counterfactual experiment tests whether reducing salience can be a useful objective for firms to pursue, and the second tests whether salience produces significant deviations from normative prescription behavior (a potential welfare concern).

Figure 3 depicts the results of our counterfactual experiments. Each bar represents the mean predicted market share for two-inhaler treatments, Seretide and Symbicort. Each of the three blocks represents one of the scenarios we compare. The first thing to note is that if AstraZeneca would have been able to eliminate salience for Symbicort (second block in Figure 3), it would have significantly increased its market share. The share of Symbicort increased, on average, by 8.5 percentage points (from 0.279 to 0.364) with 99.6% of the simulations showing an increase in market share.¹⁰ This significant increase in Symbicort’s share was mainly achieved at the expense of older two-inhaler alternatives, which lost an average of 5 percentage points (from 0.284 to 0.234) with more than 98% of the simulations resulting in a decrease of these treatments’ share. Hence, if a company alone is able to eliminate, or at least reduce, salience effects, it can reap significant market benefits. Moreover, with an additional counterfactual experiment we find that a reduction of 50% in salience achieved about one-third of the total market share effect of a full elimination

¹⁰ Note that we compare the realizations of the market shares for the two models that are based on the same set of realizations for the patient quality matches, feedback signals, etc.

Figure 3 Mean Predicted Market Shares With and Without Salience

of salience. Thus, in §9 we discuss possible salience-reducing strategies to achieve such a goal.

Finally, the shares predicted by a model with salience set to zero for all brands show that, in general, newer treatments benefit from salience elimination. The share of two-inhaler treatments decreased, on average, by 6 percentage points (from 0.284 to 0.224), with 99% of the simulations resulting in a decrease for these older treatment alternatives. In contrast, Seretide's share increased, on average, by 1.5 percentage points (from 0.437 to 0.453). We observed increases in 74% of the simulations. The prescription share of the newest entrant—Symbicort—increased 4.5 percentage points (from 0.279 to 0.324), with 89% of the simulations showing an increase.

These results indicate that in the market we study, the prevalence of salience effects in physician learning resulted in systematic changes in prescription shares, potentially with an associated welfare loss: salience (of the feedback) of switching patients slows physician learning and significantly delays the adoption of newer treatments in favor of older treatments.

8. Additional Analyses on Salience

Realizing that some concerns may persist regarding the psychological process behind the treatment-switching effect we document, we conducted additional analyses to test the robustness of our salience interpretation. We have run a survey among 156 GPs and asked these physicians to rate the importance of different drivers on their decision to prescribe an older or a newer treatment. Specifically, we compared our salience explanation with competing psychological explanations (fear about the new treatment's side effects, need to justify the decision, or potential regret). Salience was rated as significantly higher, confirming our expectations.

Another possible concern is that, in addition to salience, the treatment-switching effect may be driven by a correlated unobservable such as detailing. If a physician's decision to switch patients away from a certain treatment and the choice of the new treatment are driven by detailing, and if detailing also alters long-term market shares, our salience estimate could be inflated. We used two strategies to alleviate this concern.¹¹ First, in the physician survey we conducted, we also included detailing as a possible driver of prescription choices. Detailing was rated as significantly less important than salience by the 156 physicians. Second, please note that—if we consider a switch from treatment A to treatment B—salience predicts a penalty effect in the utility of treatment A, whereas detailing predicts a bonus effect in the utility of treatment B. Hence, we estimated a pure Bayesian learning model where we allow the number of switch-outs and the number of switch-ins to affect each treatment's utility. We find that in more than 95% of the draws, the (negative) effect of switch-outs is substantially stronger than the (positive) effect of switch-ins. This implies that switches affect the treatment that was abandoned most, in line with our salience interpretation.

9. Managerial and Public Policy Implications

Two major findings emerged from our counterfactual experiments. First, a firm that is able to eliminate or reduce salience for its treatment can gain an important competitive advantage. Considering that Symbicort had global sales of \$3,918 million in the period 2001–2006, the significant increase of 8.5 percentage

¹¹ We thank the review team for bringing this issue to our attention and suggesting strategies to deal with this problem.

points in its prescription share when its salience was set to zero would have represented a gain of \$333 million in sales. Second, setting salience to zero for all treatments confirmed that salience delays physician adoption of new treatments, suggesting a potential patient welfare loss. But what can one do to reduce salience effects?

Prior research has demonstrated that several cognitive debiasing strategies can effectively reduce biases like salience (Arkes 1991, Bradley 2005, Croskerry 2003). In the case of salience, there are at least three possible debiasing strategies we can think of. First, psychological effects, like salience, often exert their influence on judgment because people are unaware of their impact on judgments and decisions. Thus, increasing physician awareness about salience effects should help reduce its impact (Croskerry 2003). Second, prescription support systems that decrease physicians' reliance on memory in their decisions should help reduce salience effects (e.g., Bradley 2005, Croskerry 2003). Third, refreshing physicians' knowledge about the use of Bayes' rule should also help reduce salience effects (Hall 2002, Nisbett et al. 1983).

A second type of salience-reducing strategy involves streamlining marketing actions early in a treatment's life cycle. For example, firms could invest in innovations aimed at reducing patient heterogeneity in a new treatment's quality. If a firm is able to reduce patient heterogeneity for a new treatment—for instance, through new product development efforts aimed at reducing quality dispersion—it benefits both from a direct and from an indirect increase in the new treatment's utility. The former effect occurs because treatment heterogeneity increases the uncertainty about the quality of the new treatment, whereas the latter effect occurs because lower treatment heterogeneity reduces switching which, in turn, reduces salience effects.

Finally, our model also suggests that a controlled roll-out of a new drug may help speed up its adoption. Instead of aggressively targeting all patients simultaneously, firms may be better off by helping physicians to more accurately target new therapies to the patients who will most likely benefit from them. Such a strategy would represent a win-win situation whereby physicians avoid prescribing the new therapy to patients that lie on the lower tail of the quality distribution and, as a consequence, avoid undesirable switches, reducing salience effects.

10. Alternative Applications of Our Model in Marketing Science

Beyond the phenomenon studied in this paper, it is possible to adapt the model we specify to study other behavioral regularities that can be of interest to marketing scientists. First, our model can be

adapted to test—using scanner panel data—whether there is empirical evidence for positive (or negative) spillovers among different elements of a new brand's marketing mix. For instance, one could expect advertising messages to become more influential in consumer learning when the product is also featured or on display. Respecifying our salience dummy to account for these interactions would allow a researcher to quantify such spillover effects.

Second, our model can also be used to quantify the disproportionate weight given to word of mouth by dissatisfied versus satisfied consumers (e.g., Goldenberg et al. 2007). Following this route would require data on consumers' (i) purchase histories, (ii) social network, and (iii) satisfaction. Our salience parameter could then be used to quantify the extra weight given to feedback signals from dissatisfied peers in consumer learning, a metric for the disproportionate influence of unfavorable information (Mizerski 1982).

Third, our model can also be used to model confirmatory bias, i.e., consumers' tendency to pay more attention and weight more heavily information that confirms their prior beliefs (see also Boulding et al. 1999, Mehta et al. 2008). Our specification allows a researcher to model confirmatory bias directly through the weight they give to different consumption signals. For example, if we respecify the switch dummy in our model as a dummy indicating whether a certain signal confirms a consumer's prior expectation, then positive estimates for $\omega_{0,i}$ and $\omega_{\infty,i}$ can be used to directly quantify the magnitude of confirmatory bias.

11. Conclusion

In this paper, we show that salience interferes with physician learning. Patients that switch to alternative treatments are more influential during a physician's quality belief formation than patients that continue their therapy. We extend the Bayesian learning model to account for these salience effects. To the best of our knowledge, this is the first paper to uncover salience effects in physician learning using actual data on physicians' prescription choices for real patients. We find that feedback from switching patients receives between 7 and 10 times more weight in physician learning than feedback from other patients. Our finding is in line with experimental evidence that suggests that physicians are prone to use cognitive shortcuts like availability, representativeness, and anchoring and adjustment. Salience results in slower physician learning about the quality of new treatments, delaying adoption. Consequently, reducing salience effects ahead of or to a greater extent than competition, all else equal, may be very beneficial for firms that market new treatments (in our case, AstraZeneca with

Symbicort). Also, public policy officials may find the reduction of salience effects a worthwhile goal, as it represents a welfare loss. We have discussed how cognitive debiasing strategies and marketing actions may reduce salience.

11.1. Limitations and Directions for Future Research

A first limitation is our interpretation of the treatment-switching effect we quantify as a salience effect. The data we use allow us to establish that the feedback of patients who switch treatments receives significantly more weight in physicians' belief formation about the quality of a new treatment than Bayesian updating predicts. We used robust findings from psychology and medical decision-making theory and discussed additional self-reported data and analyses that reinforced our confidence that salience effects drive this treatment-switching effect. Nevertheless, it would be interesting if future research, possibly using laboratory experiments, could establish that the psychological process that underlies the treatment-switching effect we document is indeed the salience of the feedback from switching patients.

Second, we model learning solely through patient feedback. The context we have chosen (single-physician practices in a geographical market that has strict regulations on pharmaceutical marketing) limits the impact of alternative sources of information (like word of mouth and direct-to-consumer advertising). Still, if physicians' decisions to switch patients away from a certain treatment and the choice of the new treatment are driven by unobserved detailing or advertising, and if detailing also alters long-term market shares, our salience estimate could be inflated because of the well-known issue of correlated unobservables. To alleviate these concerns, we have conducted additional analyses showing that salience is significantly more likely to be the driver of the treatment-switching effect we document. It would be valuable if future studies would examine the potential for informative marketing to reduce salience effects.

Third, we assume that, at each encounter, the visiting patient only provides feedback about the last treatment she or he has been prescribed. We do not expect this assumption to introduce bias in our estimates. Still, modeling physician learning from multivariate patient feedback could allow researchers to better understand how patient and physician perceptions about different treatment alternatives interact with each other. Operationalization of such a model would require either additional data (e.g., survey data on which treatments, and what aspects of the treatment, were discussed in a certain encounter) or additional assumptions (e.g., structurally model the amount of feedback allocated to each of the

treatments a patient has previously tried). This is a very promising area for future research on consumer learning.

Finally, although we control for unobserved heterogeneity both at the patient and physician levels, observed heterogeneity could also be explored by introducing patient and physician characteristics explicitly in the model specification. Modeling across-consumer learning effects and quantifying which consumers are more influential are other areas that deserve future study.

Overall, this study confirms the usefulness of quasi-Bayesian learning models. Although such models come at the cost of increased complexity, they allow for the integration of the robust insight that human decision makers often deviate from normative rules in predictable ways into the well-established normative Bayesian learning framework.

12. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://mktsci.pubs.informs.org/>.

Acknowledgments

The authors are indebted to Arnold Vulto and António Martins Coelho for their help in interpreting medical aspects of the data. The authors thank Benedict Dellaert, Dennis Fok, Eelco Kappe, Vardit Lansdman, Mitchell Lovett, Puneet Manchanda, and Christophe Van den Bulte for their helpful comments and suggestions. The usual disclaimer applies.

References

- Arkes, H. R. 1991. Costs and benefits of judgment errors: Implications for debiasing. *Psych. Bull.* **110**(3) 486–498.
- Baumeister, R. F., E. Bratslavsky, C. Finkenauer, K. D. Vohs. 2001. Bad is stronger than good. *Rev. General Psych.* **5**(4) 323–370.
- Bornstein, B. H., A. C. Emler, G. B. Chapman. 1999. Rationality in medical treatment decisions: Is there a sunk-cost effect? *Soc. Sci. Med.* **49**(2) 215–222.
- Boulding, W., A. Kalra, R. Staelin. 1999. The quality double whammy. *Marketing Sci.* **18**(4) 463–484.
- Bradley, C. P. 2005. Commentary: Can we avoid bias? *British Medical J.* **330**(April) 784.
- Brewin, T. 1994. Primum non nocere? *Lancet* **344**(8935) 1487–1488.
- Cacioppo, J. T., W. L. Gardner. 1999. Emotion. *Annual Rev. Psych.* **50**(February) 191–214.
- Calverley, P. M., J. A. Anderson, B. Celli, G. T. Ferguson, C. Jenkins, P. W. Jones, J. C. Yates, J. Vestbo. 2007. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. *New Engl. J. Med.* **356**(8) 775–789.
- Camerer, C. F., G. Loewenstein. 2004. *Behavioral Economics: Past, Present, Future*. Princeton University Press, Princeton, NJ.
- Casella, G., E. I. George. 1992. Explaining the Gibbs sampler. *Amer. Statistician* **46**(3) 167–174.
- Chan, T., C. Narasimhan, Y. Xie. 2010. An empirical model of physician learning on treatment effectiveness and side-effects. Working paper, Washington University in St. Louis, St. Louis.

- Chib, S., E. Greenberg. 1995. Understanding the Metropolitan-Hastings algorithm. *Amer. Statistician* 49(4) 327–335.
- Ching, A., M. Ishihara. 2010. The effects of detailing on prescribing decisions under quality uncertainty. *Quant. Marketing Econom.* 8(2) 123–165.
- Chintagunta, P. K., R. Jiang, G. Z. Jin. 2009. Information, learning, and drug diffusion: The case of Cox-2 inhibitors. *Quant. Marketing Econom.* 7(4) 399–443.
- Coscelli, A. 2000. The importance of doctors' and patients' preferences in the prescription decision. *J. Indust. Econom.* 48(3) 349–369.
- Coscelli, A., M. Shum. 2004. An empirical model of learning and patient spillovers in new drug entry. *J. Econometrics* 122(2) 213–246.
- Crawford, G. S., M. Shum. 2005. Uncertainty and learning in pharmaceutical demand. *Econometrica* 73(4) 1137–1173.
- Croskerry, P. 2002. Achieving quality in clinical decision making: Cognitive strategies and detection of bias. *Acad. Emergency Med.* 9(11) 1184–1204.
- Croskerry, P. 2003. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad. Med.* 78(8) 775–780.
- Datamonitor. 2008. Forecast insight: Asthma/COPD—Little long-term future for long-acting single-agent inhalers? Report DMHC2439, Datamonitor, London.
- DeGroot, M. H. 1970. *Optimal Statistical Decisions*. John Wiley & Sons, Hoboken, NJ.
- Elstein, A. S., A. Schwartz. 2002. Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *British Medical J.* 324(7339) 729–732.
- Epstein, L. G. 2006. An axiomatic model of non-Bayesian updating. *Rev. Econom. Stud.* 73(2) 413–436.
- Erdem, T., M. P. Keane. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Sci.* 15(1) 1–20.
- Estrada, C. A., A. M. Isen, M. J. Young. 1997. Positive affect facilitates integration of information and decreases anchoring in reasoning among physicians. *Organ. Behav. Human Decision Processes* 72(1) 117–135.
- Fiscella, K., P. Franks, J. Zwanziger, C. Mooney, M. Sorbero, G. C. Williams. 2000. Risk aversion and costs: A comparison of family physicians and general internists. *J. Family Practice* 49(1) 12–17.
- Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith, eds. *Bayesian Statistics*, Vol. 4. Oxford University Press, Oxford, UK, 169–193.
- Global Initiative for Asthma (GINA). 2008. Global strategy for asthma management and prevention. Report, GINA, <http://www.ginasthma.com/Guidelineitem.asp??/1=2&/2=1&intId=60>.
- Goldenberg, J., B. Libai, S. Moldovan, E. Muller. 2007. The NPV of bad news. *Internat. J. Res. Marketing* 24(3) 186–200.
- Griffin, D., A. Tversky. 1992. The weighing of evidence and the determinants of confidence. *Cognitive Psych.* 24(3) 411–435.
- Hall, K. H. 2002. Reviewing intuitive decision-making and uncertainty: The implications for medical education. *Medical Ed.* 36(3) 216–224.
- Häubl, G., B. G. C. Dellaert, B. Donkers. 2010. Tunnel vision: Local behavioral influences on consumer decisions in product search. *Marketing Sci.* 29(3) 438–455.
- Ho, T. H., N. Lim, C. F. Camerer. 2006. Modeling the psychology of consumer and firm behavior with behavioral economics. *J. Marketing Res.* 43(3) 307–331.
- Kass, R. E., A. E. Raftery. 1995. Bayes factors. *J. Amer. Statist. Assoc.* 90(430) 773–795.
- Klein, J. G. 2005. Five pitfalls in decisions about diagnosis and prescribing. *British Medical J.* 330(7494) 781–783.
- Marks, N. A., P. W. Ind. 2005. Novel therapies in asthma: Long-acting β_2 -agonists/inhaled corticosteroids. P. G. Gibson, M. Abramson, U. Costabel, M. Hensley, J. Volmink, R. Wood-Baker, eds. *Evidence-Based Respiratory Medicine*. Blackwell Publishing, Malden, MA, 217–230.
- McCulloch, R., P. E. Rossi. 1994. An exact likelihood analysis of the multinomial probit model. *J. Econometrics* 64(1–2) 207–240.
- Medin, D. L., M. W. Altom, S. M. Edelson, D. Freko. 1982. Correlated symptoms and simulated medical classification. *J. Experiment. Psych.: Learn., Memory, Cognition* 8(1) 37–50.
- Mehta, N., X. J. Chen, O. Narasimhan. 2008. Informing, transforming, and persuading: Disentangling the multiple effects of advertising on brand choice decisions. *Marketing Sci.* 27(3) 334–355.
- Mehta, N., S. Rajiv, K. Srinivasan. 2004. Role of forgetting in memory-based choice decisions: A structural model. *Quant. Marketing Econom.* 2(2) 107–140.
- Mizerski, R. W. 1982. An attribution explanation of the disproportionate influence of unfavorable information. *J. Consumer Res.* 9(3) 301–310.
- Narasimhan, C., C. He, E. T. Anderson, L. Brenner, P. Desai, D. Kuksov, P. Messinger et al. 2005. Incorporating behavioral anomalies in strategic models. *Marketing Lett.* 16(3/4) 361–373.
- Narayanan, S., P. Manchanda. 2009. Heterogeneous learning and the targeting of marketing communication for new products. *Marketing Sci.* 28(3) 424–441.
- Narayanan, S., P. Manchanda, P. K. Chintagunta. 2005. Temporal differences in the role of marketing communication in new product categories. *J. Marketing Res.* 42(3) 278–290.
- Nisbett, R. E., D. H. Krantz, C. Jepson, Z. Kunda. 1983. The use of statistical heuristics in everyday inductive reasoning. *Psych. Rev.* 90(4) 339–363.
- Oliver, R. L. 1993. Cognitive, affective, and attribute bases of the satisfaction response. *J. Consumer Res.* 20(3) 418–430.
- Poses, R. M., M. Anthony. 1991. Availability, wishful thinking, and physicians' diagnostic judgments for patients with suspected bacteremia. *Medical Decision Making* 11(3) 159–168.
- Rabin, M., J. L. Schrag. 1999. First impressions matter: A model of confirmatory bias. *Quart. J. Econom.* 114(1) 37–82.
- Raftery, A. E., S. M. Lewis. 1992. How many iterations in the Gibbs sampler? J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith, eds. *Bayesian Statistics*, Vol. 4. Oxford University Press, Oxford, UK, 763–773.
- Redelmeier, D. A. 2005. The cognitive psychology of missed diagnoses. *Ann. Internal Med.* 142(2) 115–120.
- Roberts, J. H., G. L. Urban. 1988. Modeling multiattribute utility, risk, and belief dynamics for new consumer durable brand choice. *Management Sci.* 34(2) 167–185.
- Rossi, P. E., G. M. Allenby, R. McCulloch. 2005. *Bayesian Statistics and Marketing*. John Wiley & Sons, West Sussex, UK.
- Rozin, P., E. B. Royzman. 2001. Negativity bias, negativity dominance, and contagion. *Personality Soc. Psych. Rev.* 5(4) 296–320.
- Stremersch, S. 2008. Health and marketing: The emergence of a new field of research. *Internat. J. Res. Marketing* 25(4) 229–233.
- Stremersch, S., W. Van Dyck. 2009. Marketing of the life sciences: A new framework and research agenda for a nascent field. *J. Marketing* 73(July) 4–30.
- Swait, J., J. Louviere. 1993. The role of the scale parameter in the estimation and comparison of multinomial logit models. *J. Marketing Res.* 30(3) 305–314.

- Tanner, M. A., W. H. Wong. 1987. The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* **82**(398) 528–540.
- Taylor, S. E. 1991. Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psych. Bull.* **110**(1) 67–85.
- Tversky, A., D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* **185**(4157) 1124–1131.
- Venkataraman, S., S. Stremersch. 2007. The debate on influencing doctor's decisions: Are drug characteristics the missing link? *Management Sci.* **53**(11) 1688–1701.
- Vlug, A. E., J. van der Lei, B. M. Mosseveld, M. A. van Wijk, P. D. van der Linden, M. C. Sturkenboom, J. H. van Bommel. 1999. Postmarketing surveillance based on electronic patient records: The IPCI project. *Methods Inform. Med.* **38**(4–5) 339–344.