# Analytics for Customer Engagement

Bijmolt, T.H.A.; Leeflang, P.S.H.; Block, F.; Eisenbeiss, M.; Hardie, B.G.S.; Lemmens, Aurélie; Saffert, P.

*Published in:*
Journal of Service Research

*Publication date:*
2010

Journal of Service Research
13(3) 341-356

DOI: 10.1177/1094670510375603
http://jsr.sagepub.com

# Analytics for Customer Engagement

**Tammo H. A. Bijmolt[1], Peter S. H. Leeflang[1], Frank Block[2], Maik Eisenbeiss[3], Bruce G. S. Hardie[4], Aurélie Lemmens[5], and Peter Saffert[3]**

**Abstract**

In this article, we discuss the state of the art of models for customer engagement and the problems that are inherent to calibrating and implementing these models. The authors first provide an overview of the data available for customer analytics and discuss recent developments. Next, the authors discuss the models used for studying customer engagement, where they distinguish the following stages: customer acquisition, customer development, and customer retention. Finally, they discuss several organizational issues of analytics for customer engagement, which constitute barriers for introducing analytics for customer engagement.

**Keywords**

analytic models, customer equity, customer management, data mining, decision trees, probability models, word of mouth

## Introduction

The classic view is that the customer is exogenous to the firm and is the passive recipient of the firm's active value creation efforts, and values created in "the factory" (Deshpandé 1983). A different perspective is now emerging, namely, that customers can cocreate value, cocreate competitive strategy, collaborate in the firm's innovation process, and become endogenous to the firm. Central in this new view is the concept of *customer engagement*, defined as the behavioral manifestation from a customer toward a brand or a firm which goes beyond purchase behavior (Van Doorn et al. 2010). This behavioral manifestation may affect the brand or firm and its constituents in ways other than purchase such as word-of-mouth (WOM) referrals, participation in the firm's activities, suggestions for service improvements, customer voice, participation in brand communities, or revenge activities. As a consequence, the relation between a company and its customers becomes closer, more selective and may become so familiar that even the term intimacy is used (Treacy and Wiersema 1993). Given that not all customers appreciate that intimacy, firms will be confronted with engaged-prone customers and "other customers."

Customer engagement is connected to customer value management (Verhoef, van Doorn, and Dorotic 2007) through its objective, namely, to maximize the value of a firm's customer base. However, in customer value management, the value of a customer is generally linked to *direct* customer outcomes such as its current and future *transactions* with the firm. In contrast, customer engagement (additionally) includes behavioral manifestations of a customer with a rather *indirect* impact on firm performance. In particular, we distinguish between three general manifestations of customer engagement: WOM, customer cocreation, and complaining behavior; all of which affecting the brand or firm in ways other than purchase.

Neglecting behavioral manifestations of this kind can lead to a highly biased perception of a customer's contribution to a firm. For example, von Wangenheim and Bayón (2007) find that the lack to incorporate WOM in the customer lifetime value (CLV) calculation could lead to an underestimation of the CLV by up to 40%. Thus, it seems essential to establish measures and models accounting for key behavioral manifestations of customer engagement. Despite its relevance, this issue has actually been scarcely researched in literature.

The use of customer analytics may have a positive impact on firm performance (Hoekstra and Verhoef 2010). With customer analytics, we mean the extensive use of data and models and fact-based management to drive decisions and actions, where data and models are defined at the individual customer level (based on Davenport and Harris 2007, p. 7). However, most analytical models that have been developed focus on customer transactions. Despite this rather narrow perspective, these more "traditional" and well-established models provide a promising starting point for discussing how customer engagement reflecting behavioral manifestations other than purchase may be modeled appropriately. Building on this link, the main objective of

[1] Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands
[2] FinScore S.A., Renens, Switzerland
[3] University of Cologne, Cologne, Germany
[4] London Business School, London, United Kingdom
[5] Erasmus University Rotterdam, The Netherlands

**Corresponding Author:**
Tammo H. A. Bijmolt, Department of Marketing, Faculty of Economics and Business, University of Groningen, PO Box 800, 9700 AV, Groningen, The Netherlands
Email: t.h.a.bijmolt@rug.nl

this article is to discuss how existing knowledge and modeling approaches from the transaction area may be leveraged for model building in a customer engagement context. Moreover, taking into consideration the increasing ease to quickly interact online and the resulting customer engagement opportunities (e.g., customer cocreation; Hoyer et al. 2010), a closer look at the capabilities of analytical methods to deal with large data sets and their computing time is also important. In particular, key aims of this article are

1. To review opportunities and organizational aspects with respect to data collection for customer engagement;
2. To give a brief overview of "traditional" models dealing with customer transactions and to subsequently discuss how key behavioral manifestations of customer engagement (WOM, cocreation, and complaining behavior) can be included in these models;
3. To discuss problems that are apparent in marketing practice and constitute barriers for introducing analytics for customer engagement.

The order of discussion is as follows. First, we provide an overview of the data available for customer analytics, which is followed by an overview of models for customer engagement. Customer engagement may be generated in different stages of the customer life cycle: customer acquisition, customer development (growth), and customer retention (churn and win-back). We discuss models that can be used as supporting tools for each of these stages. In the final section, we discuss the gaps between the state-of-the-art tools that marketing scientists create and what marketing managers actually use and suggest directions for future developments closing these gaps.

## Data for Customer Analytics

Twenty years ago, marketing scientists were starting to come to grips with the new single-source scanner data sets made possible by the widespread adoption of the Universal Product Code (UPC) and the associated scanners at the supermarket checkout. For those researchers that worked with bimonthly audit data from ACNielsen, these data sets seemed huge. How times have changed. Developments in data collection and data storage technologies mean that marketing databases have proliferated and grown in both size and complexity and new sources of data have emerged. These pose a number of challenges that are being largely ignored by marketing scientists.

When faced with a huge data set, the obvious approach is to work with a sample of the data set. One problem with this is that inaccuracy is introduced by sampling variance, which may not be acceptable in situations where the detection of small differences in a small subset of the overall population is very important (e.g., fraud detection). One solution to this problem is data squashing (DuMouchel 2002). The idea is to create a smaller version of the data set that mimics the original, which is then used for model development. Such developments should be explored by marketing scientists.

Another approach is to aggregate the data in some manner and try to fit the models to these aggregated data. For example, Fader, Hardie, and Jerath (2007) explore how the Pareto/negative binomial distribution (NBD) model can be fitted to a series of histograms of the number of purchases made per year by a cohort of customers. In some settings, it may be possible to develop parametric models of the phenomenon of interest where the sufficient statistics required for model estimation are simple data summaries (Albuquerque and Bronnenberg 2009).

Large data sets create other challenges for the marketing scientist. The past 15 years have seen Bayesian statistical methods evolve from being a topic of intellectual curiosity to an essential component of any marketing scientist's toolkit (Rossi, Allenby, and McCulloch 2005). Central to this shift was the development of Markov Chain Monte Carlo (MCMC) methods that involve thousands, if not millions, of passes through the data. As a result, it is typically not feasible to use traditional Bayesian methods on massive data sets because of computational resource constraints. There is an emerging literature on the scaling-up of Bayesian methods to massive data sets. For example, Balakrishnan and Madigany (2006) propose an approach by which a rigorous Bayesian analysis is first performed on a small portion of the data set and then adapts those calculations for the whole data set by making a single pass through the remaining observations.

Next to the increased size of databases, an important issue is that organizations think more about operating in "real time" (Goldenberg 2008). Hence, another data "reality" is that of processing streams of data in real time. A number of statisticians are working in this area (e.g., Lambert, Pinheiro, and Sun 2001; McDermott et al. 2007) and those marketing scientists interested in customer analytics may wish to track this research. While such real-time analysis is vital in telecommunications, financial services, and online settings, it may not be necessary for other customer analytics settings. There is a need to think about the "need for speed" in various analytics activities supporting management of customer engagement.

The same information technology developments that have lead to the massive growth in customer data have also lowered the costs of some traditional data sources while facilitating new data sources. For example, online surveys tools make it extremely easy for firms to survey their own customers. As such, work on the marrying attitudinal data collected via surveys and transaction data (e.g., Kamakura et al. 2003) becomes even more important.

Finally, a substantial part of customer behavior occurs in an online setting, resulting in new sources of data for studying customer engagement. For example, the emergence of social media—the "group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation and exchange of User-Generated Content" (Kaplan and Haenlein 2010) forms an important development for the customer–firm relationship.

### *Future Model Directions for Customer Engagement*

The challenge will be how to extract insights from the new huge and complex databases and how to incorporate them into our models for customer engagement. Marketing scientists interested in working in this area will need to learn about large-scale text mining techniques or collaborate with researchers from other disciplines that already possess these skills.

## Models for Customer Engagement

In the literature, numerous analytical models have been developed to study customer–firm relationships (models for customer value management). Despite a clear focus on purchase-related outcomes, these more "traditional" and well-established models provide a promising starting point for discussing how customer engagement reflecting behavioral manifestations other than purchase may be modeled appropriately. Building on this link, we first review the broad range of models and techniques that have been used in the context of customer value management. Subsequently, we discuss how key behavioral manifestations of customer engagement, that is, WOM, cocreation, and complaining behavior, may be included in analytical models of this kind while referring to both recent developments and future research directions. Since the various manifestations of customer engagement may be generated in different stages of the customer life cycle, we structure our discussion along three key stages: customer acquisition, customer development, and customer retention (Kamakura et al. 2005; Reinartz, Krafft, and Hoyer 2004). Tables 1 and 2 provide an overview of the discussed analytical models for customer transactions and customer engagement, respectively.

### *Models for Customer Acquisition*

*Customer Selection.* The initial goal of customer acquisition is to select the "right" prospects for the acquisition campaign. Depending on the objective function, a "right" prospect can be someone with maximum response likelihood, maximum purchase probabilities/levels or, as most in line with the customer relationship management (CRM) principles, maximum expected CLV.

Historically, the most frequently used selection technique has been the Recency, Frequency, and Monetary value (RFM) model. Building on the assumption that the "right" customer in the future looks a lot like the "right" customer in the past, the traditional RFM modeling approach creates groups of customers based on their RFM characteristics of prior purchases and then assigns probabilities or "scores" to each group in accordance with its differential response behavior. Marketing programs such as mailing campaigns are then prioritized based on the scores of different RFM groups (Gupta et al. 2006). Extensions of the RFM scoring approach define the customer groups using other behavioral (non-RFM) or sociodemographic variables. Statistical models that are frequently used include Automatic Interaction Detection (AID) and Chi-square Automatic Interaction Detection (CHAID) selection techniques (David Shepard Associates 1999), parametric regression-based (scoring) models (Malthouse and Blattberg 2005), discriminant analysis, and log-linear models (LLM). In response to the predetermined functional form and restrictive model assumptions of many (parametric) scoring models, more flexible techniques such as semiparametric regression models (Bult and Wansbeek 1995) or neural networks (Baesens et al. 2002; Malthouse and Blattberg 2005) have been proposed. Baesens et al. (2002) demonstrate in an empirical study that, from a predictive performance perspective, (Bayesian) neural network analysis performs significantly better than logistic regression and discriminant analysis classifiers.

Although scoring models seem highly appealing and are easy to use for customer selection, they have several known shortcomings. First, scoring models predict behavior in the next period only. However, estimating CLV for example not only requires information from the next period but also from periods thereafter (Fader, Hardie, and Lee 2005b; Gupta et al. 2006). Second, at least two data sets are required for estimation and validation. When the same data are used for model calibration as for model validation and comparison, it is possible that the model explains noise instead of the underlying relationship, a problem known as overfitting. As a result, the model is wrongly assumed to perform better than it will in practice. Overfitting is particularly a problem with very flexible modeling techniques such as neural networks, CHAID, or semiparametric regression (Malthouse 2001). Finally, results of a firm's past marketing activities on consumers' past behavior are not accounted for in scoring models (Fader, Hardie, and Lee 2005).

One class of models that overcomes many of these limitations consists of probability models (for a comprehensive review see Fader and Hardie 2009). The key underlying idea of the probability modeling approach is that observed behavior is a function of an individual's latent behavioral characteristics. After fitting a probability model to the data, inferences about an individual's latent characteristics can be made, given his observed behavior, which, in turn, can be used for predicting future behavior (Fader and Hardie 2009).

All models for customer selection discussed so far are suited for "classical" RFM data, that is, purchase data collected in physical stores. However, one of the main challenges of models for customer acquisition is that the transaction history is not available for prospects. Hence, the researcher is left with less informative variables such as demographics and psychographics for profiling the top tier customers and identifying prospects that resemble these top tier customers. The models developed in the online marketing literature, for example, Moe and Fader (2004), provide insights on how to deal with this challenge by exploiting clickstream data or data on other non-purchase behavior.

Just as purchase data can be collected in physical stores, it can also be collected in virtual stores, with the only difference that the data set of the virtual store entails more information (Montgomery and Smith 2009; for an overview of clickstream data analysis in Marketing, see Bucklin and

**Table 1.** Analytical Models Focusing on Customer Transactions

| | Models for | | | |
|---|---|---|---|---|
| | Customer Acquisition | | Customer Development | Customer Retention |
| Type of data | – RFM data<br>– Customer characteristics (e.g., demographics)<br>– Company-interaction variables (e.g., marketing actions)<br>– Clickstream data | | – Cross-sectional and longitudinal customer data in contractual or noncontractual setting | – Cross-sectional customer data in contractual or non-contractual setting<br>– Satisfaction surveys |
| Type of methods | *Customer selection* | *Managing customer acquisitions* | | |
| | Parametric (scoring) models<br>– RFM scoring (Gupta et al. 2006)<br>– CHAID (David Shepard Associates 1999)<br>– Linear regression model (Bauer 1988; Berger and Magliozzi 1992; Malthouse and Blattberg 2005)<br>– Sequential probit model (Sismeiro and Bucklin 2004)<br>– Latent class probit model (Vroomen, Donkers, Verhoef, and Franses 2005) | Parametric models<br>– Logit/Probit model (Hansotia and Wang 1997; Lewis 2005; Reinartz, Thomas, and Kumar 2005; Verhoef and Donkers 2005)<br>– Tobit model (Hansotia and Wang 1997; Lewis 2006)<br>– Hazard model (Thomas, Blattberg, and Fox 2004)<br>– Generalized gamma model (Venkatesan and Kumar 2004)<br>– Hierarchical Bayesian model (Ansari and Mela 2003)<br>– Poisson count model (Anderson and Simester 2004) | Parametric models<br>– Regression model (Bowman and Narayandas 2004; Venkatesan and Kumar 2004)<br>– Logistic regression model (Knott, Hayes, and Neslin 2002)<br>– Multinomial logit model (Knott, Hayes, and Neslin 2002)<br>– Discriminant analysis (Knott, Hayes, and Neslin 2002)<br>– Multivariate probit model (Li, Sun, and Wilcox 2005)<br>– Tobit type 2 model (Ansari, Mela, and Neslin 2008)<br>– Hazard model (Venkatesan, Kumar, and Ravishankar 2007) | Parametric (binary prediction) models<br>– Logit/Probit model (Donkers, Verhoef, and de Jong 2007)<br>Nonparametric (binary prediction) models<br>– Neural networks (Hung, Yen, and Wang 2006)<br>– Support-vector machine (Coussement and Van den Poel 2008)<br>– Aggregating methods: Bagging, Boosting (Lemmens and Croux 2006) |
| | Semi-/Nonparametric (scoring) models<br>– Semiparametric logit model (Bult and Wansbeek 1995; Malthouse 2001)<br>– Neural networks (Baesens et al. 2002; Malthouse and Blattberg 2005) | Data/web usage mining<br>– Transaction/usage clustering (Mobasher, Cooley, and Srivastava 2000)<br>– Association rule discovery (Mobasher, Cooley, and Srivastava 2000; Mobasher et al. 2001)<br>– Fuzzy inference engine (Nasraoui and Petenes 2003)<br>– UBB Mining (Ting, Kimble, and Kudenko 2005) | Semi-/Nonparametric models<br>Neural networks (Knott, Hayes, and Neslin 2002)<br>– Persistence models<br>– Autoregressive discrete choice model (Yang and Allenby 2003) | Parametric (duration) models<br>Hazard model (Schweidel, Fader, and Bradlow 2008) |
| | Probability models<br>– Pareto/NBD model (Abe 2009; Fader, Hardie, and Lee 2005b; Schmittlein, Morrison, and Colombo 1987)<br>– Beta-geometric/NBD model (Fader, Hardie, and Lee 2005a; Fader, Hardie; and Shang (2010)<br>– Individual-level probability model (Moe and Fader 2004) | | Probability models<br>– Markov decision processes (Tirenni et al. 2007)<br>– Hidden Markov model (Netzer, Lattin, and Srinivasan 2008)<br>– Stochastic choice model (Bodapati 2008)<br>Other models<br>– Matching methods (Hitt and Frei 2002; Gensler et al. 2010) | Probability models<br>– Pareto/NBD model (Schmittlein, Morrison, and Colombo 1987)<br>– Shifted-beta-geometric (Fader and Hardie 2007) |

*Note.* NBD = negative binomial distribution.

**Table 2.** Analytical Models Focusing on Customer Engagement

| | Models for | | | |
|---|---|---|---|---|
| | Customer Acquisition | | Customer Development | Customer Retention |
| Type of data | Clickstream data<br>Word-of-mouth data | | Brand community and information website data<br>Word-of-mouth data | Loyalty program data<br>Social network data<br>Time-series data |
| Type of methods | *Customer selection* | *Managing customer acquisitions* | | |
| | Probability models<br>– Truncated NBD model (Bowman and Narayandas 2001)<br>– Zero-inflation Poisson model (Wangenheim and Bayón 2007) | Persistence models<br>– Vector Autoregressive (VAR) model (Villanueva, Yoo, and Hanssens 2008) | Parametric models<br>– Structural equation model (Bagozzi and Dholakia 2006) | Parametric models<br>– Dependence between customer value, customer engagement and churn probability (Lemmens and Croux 2010)<br>– Churn dependence across customers (Gupta and Mela 2008) |
| | | | Persistance models<br>– VAR model (Trusov, Bucklin, and Pauwels 2009)<br>Other models<br>– Agent-based model (Goldenberg et al. 2007) | Dynamic churn models<br>– (Hidden) Markov model (Netzer, Lattin, and Srinivasan 2008; Pfeifer and Carraway 2000)<br>– Time-varying coefficient or dynamic linear model |

*Note*. NBD = negative binomial distribution.

Sismeiro 2009). In particular, in online shopping environments, it can not only be observed if, when, and what the customer purchased but also visitor movements through the store can be tracked; that is, what items visitors looked at, how long they considered their decisions, in which sequence they bought items, and so on. Clickstream data of this kind available to (online) managers have led to another stream in research focusing on the development of models that account for the particularities of these data sets and help extract as much information as possible out of them. Many of these models, at least implicitly, address customer selection issues such as modeling and predicting customers' future purchasing probability. For example, Moe and Fader (2004) offer an individual-level probability model that predicts the visits that are likely to convert to purchases. Not only does their model control for different forms of customer heterogeneity it also allows shopping behavior to evolve over time as a function of prior experiences. Contrary to Moe and Fader (2004), Sismeiro and Bucklin (2004) take individual-level sequencing information into account and propose a sequential probit model that predicts online buying by linking the purchase decision to what visitors do and to what they are exposed to while browsing a particular website. Finally, Montgomery et al. (2004) propose a dynamic multivariate probit model that allows capturing page-level movements through a website for the prediction of purchase conversion. They model clickstream data on a very disaggregate level, thereby improving the predictive power with regard to understanding which users are likely to make a purchase and which are not.

*Managing Customer Acquisitions.* Once the decision regarding which customers to focus on has occurred, the next question to be addressed is how to allocate resources among marketing variables for leveraging customer acquisition. Firms use various types of marketing activities for customer acquisition, which differ according to the communication channel through which a prospect is acquired and the message that is used to attract the prospect (Reinartz and Venkatesan 2008). At the acquisition channel level, firms can acquire customers directly (i.e., marketing induced customer acquisition) through one of the following channels: personal selling, mass media (e.g., radio and television), direct marketing channels (e.g., direct mail and telemarketing), Internet, and retail outlets (Bolton, Lemon, and Verhoef 2004). At the same time, firms can use different messages (in terms of content and design) to attract different customers. For example, messages may contain brand-related information or price-related information.

Under the notion that different acquisition channels lead to different "qualities" of customers (Lewis 2006; Villanueva, Yoo, and Hanssens 2008), researchers have modeled the effectiveness of different acquisition channels and have developed models to allocate the acquisition budget more efficiently. Most of the applied models are probability models, which incorporate covariates to explain variation in selected customer profitability metrics. For example, Verhoef and Donkers (2005) use variants of probit models to explore how retention rates and cross-selling opportunities differ among the various acquisition channels a financial services provider uses. Reinartz, Thomas, and Kumar (2005) use a probit two-stage least square model to link customer acquisition to relationship duration and profitability. Related to that, Venkatesan and Kumar (2004) develop a panel-based stochastic model that provides guidance on how much to invest in distinct communication channels.

Content-related and design-related attributes of the acquisition message are another key element of any marketing-induced acquisition campaign. Several researchers have modeled the effect of price discounts on various customer metrics and have provided models to improve price-related decision making in the context of customer acquisition management. For example, Anderson and Simester (2004) use a Poisson count model to conclude that customers acquired through catalogues with more discounted items have higher long-term value. In contrast, Lewis (2006) who uses several modeling approaches (logistic regression, accelerated failure time (AFT) hazard model, and Tobit model) found that acquisition discount depth is negatively related to repeat-buying rates and customer asset value. Besides the price-related studies, Ansari and Mela (2003) focus on the design of communications or marketing programs for customer acquisition purposes. Using clickstream data, they develop a statistical and optimization approach for customization of information on the Internet while modeling the effects of different types and different ordering of targeted e-mail messages to increase the likelihood of purchase.

*Future Model Directions for Customer Engagement.* At the heart of the above-mentioned "traditional" models for *customer selection* is an objective function used to discriminate among prospects who differ in terms of their response likelihood for a campaign or their purchase level. In the context of customer engagement, these rather purchase-related objectives can readily be replaced by behavioral manifestations other than purchase and be included in the "traditional" modeling approaches. For example, instead of predicting future purchase levels, analytical models may be aligned to predict the number of WOM referrals. Given the assumption that WOM communication positively affects revenues, firms may be interested in targeting customers with a high propensity to WOM. Bowman and Narayandas (2001) estimate two models for predicting WOM. First, using a logistic regression model, they determine whether a WOM referral is made. Then, applying a truncated-at-zero NBD model, they estimate the actual number of referrals, given that at least one referral was made. A slightly more convenient approach for predicting WOM is offered by zero-inflation models allowing a joint estimation of the binary and the count model. For example, von Wangenheim and Bayón (2007) use a zero-inflation Poisson (ZIP) model, in which the standard Poisson model is complemented by a logit model. While the logit specification determines whether a referral is made, the Poisson count model subsequently predicts the number of referrals. An appealing side aspect of this model is that it allows for different sets of independent variables predicting the binary and the Poisson model.

A promising direction for future research in this area would be to account for the long-term effects of WOM communication, since WOM behavior is likely to change over the customer life cycle. For example, von Wangenheim and Bayón (2007) reveal that "new" customers are trying to communicate the goodness of their choice more heavily to others than "older" ones. Accounting for long-term effects of this kind becomes especially critical, when WOM should be incorporated in CLV calculations.

At the acquisition channel level, firms not only acquire customers directly but also indirectly through referrals from the prospects' social network. Under the assumption that different acquisition channels lead to different "qualities" of customers (Lewis 2006), firms also need to understand in which way the fact that a customer acquired through WOM impacts its lifetime value. A suitable model for this issue has been proposed by Villanueva, Yoo, and Hanssens (2008). The authors' use of a VAR model in order to capture WOM effects of a new customer acquisition on customer equity growth. Such a VAR model belongs to a class of models commonly referred to as persistence models (Gupta et al. 2006). Persistence models can be used to assess the long-term consequences of the various acquisition tactics, including customer selection, acquisition channels, as well as acquisition message. Persistence models focus on modeling the acquisition process and its link to customer-based metrics as part of a dynamic system, which, from a methodological point of view, requires the use of time-series data.

From a strategic perspective understanding how information communicated through mass media (external influence) and then spread through WOM (internal influence) affects the process of consumer adoption has great importance for customer acquisition. Different research methodologies attempt to investigate the role and measurement of WOM:

1. There is a classic line of modeling based on Bass' (1969) well-known adoption model. This classic line of research attempts to explain how marketing mix strategies affect new product diffusions (Mahajan, Muller, and Wind 2000) and shows that WOM effectively encourages people to start using a product (Hess, Kardes, and Kim 1991). Usually, these models use aggregate data.
2. The class of agent-based simulation models is a methodology that is especially useful when the agent rules and characteristics can be defined on an individual level, when the population that adopts a new product is heterogeneous, and/or when the topology of the interactions between individuals is complex and heterogeneous. In marketing practice, these investigations have initiated strategies such as "viral" and "buzz" marketing. Examples of studies that use agent-based modeling are Goldenberg et al. (2009) and Van Eck, Jager, and Leeflang (2010).

Another promising research area to account for customer engagement in the context of acquisition models is to predict consumers' willingness/ability to engage in cocreation activities for new product development (Fuchs, Prandelli, and Schreier 2010). Since consumers often vary highly in their willingness and ability to participate in cocreation tasks (Hoyer et al. 2010), firms are becoming increasingly interested in preselection mechanisms to identify segments of consumers who might be particularly willing and able to participate (Hoffman, Kopalle, and Novak 2010). Here, customer selection models incorporating major drivers of customer willingness and ability to cocreate, such as scoring models can be applied. In addition, cocreation—especially in an online environment—is likely to produce large volumes of consumer input that often requires a firm "screening millions of ideas" (Hoyer et al. 2010). Again, in order to overcome this problem, firms may be interested in suitable preselection models. Since most of these issues are relevant for Internet-based cocreation mechanisms, the above-discussed "traditional" models for clickstream data provide a particularly suitable starting point for future research in this field.

In addition, participation levels in cocreation efforts may vary over time within a relationship with a consumer (Hoyer et al. 2010). For example, an initial motivation to cocreate may diminish over time while, in the same way, it could also intensify with relationship duration. Thus, a promising direction for future research in this area would be to account for the long-term evolution of cocreation behavior and its impact on CLV.

## Models for Customer Development

In the development stage, a CLV can be stimulated through many marketing activities. This ultimately results in growth in sales among existing customers, cross-buying, and upgrading (Verhoef, Van Doorn, and Dorotic 2007). Given that a number of excellent reviews have been recently published (Blattberg, Kim, and Neslin 2008, Chapter 25; Reinartz and Venkatesan 2008), we suffice to mention only areas in which customer analytics play an important role.

Many studies that propose models for customer development have appeared in recent years. One of the core issues for customer development is the estimation of the CLV. So far, most of the approaches rely on customer transactions (compare for example Gupta et al. 2006). One class of models tries to predict a consumer's margin directly using econometric methods. For example, Venkatesan and Kumar (2004) use a regression model based on past transactions and marketing mix variables to predict the contribution margin. Using systems of equations within an extended service-profit chain framework, Bowman and Narayandas (2004) link customer management efforts to customer profitability.

Another class of models investigates cross-selling (e.g., Kamakura 2008). Building on the idea that customers have predictable life cycles and, as a result, buy certain products before others, Li, Sun, and Wilcox (2005) model the demand for multiple products of the same provider (of banking services) over time. To this end, they use a multivariate probit model. Related to this, Knott, Hayes, and Neslin (2002) apply a logit model specification, discriminant analysis, and neural networks to predict the next product to buy. Lemon and Von Wangenheim

(2009) develop a "dynamic" model of cross-buying across loyalty program partnerships using data from European airlines. Reinartz, Thomas, and Bascoul (2008) apply a causal model to investigate the direction of the relationship between cross-buying and behavioral loyalty. They find that purchasing items for multiple categories is by and large a consequence of behavioral loyalty not an antecedent. Other approaches to cross-selling comprise recommendation systems. Bodapati (2008) uses a hierarchical Bayes method with a dual latent class structure to develop a model for firms' recommendations to customers. Netzer, Lattin, and Srinivasan (2008) propose a hidden Markov model for the transitions among latent relationship states and its effects on buying behavior. The model is intended to guide the firm's marketing decisions to alter the long-term buying behavior of its customers.

An important development for consumers is their increased opportunity to collect information and to order products from many channels. These channels include the Internet, call centers, sales forces, catalogues, retail stores, interactive television, and so on (Blattberg, Kim, and Neslin 2008). Customers not only have more opportunities to contact firms, but the number of opportunities for home delivery increased also. Companies such as Peapod and Streamline allow their customers to organize, in a customized fashion, their shopping behavior electronically already for some time. A related topic that receives also much attention is the modeling of channel migration strategies: Hitt and Frei (2002) and Gensler, Leeflang, and Skiera (2010) use matching methods to determine the effects of channel migration.

*Future model directions for customer engagement.* Accounting for customer engagement in the development stage requires understanding how behavioral manifestations such as WOM, cocreation activities, and complaining behavior impact a CLV.

With regard to WOM, Goldenberg et al. (2007) explored the effects of individual and network-level negative WOM on profits using an agent-based model. They found that the effect of negative WOM on the Net Present Value (NPV) of the firm is substantial, even when the initial number of dissatisfied customers is relatively small. Trusov, Bucklin, and Pauwels (2009) study the effect of WOM on member growth at an Internet social networking site. The authors employ a VAR modeling approach. They find that WOM-elasticities are approximately 20 times higher than that of marketing events and 10 times that of media appearances.

Brand communities can create value among networked firm-facing actors, as such, the active management and stimulation to cocreate is another important task. Case studies show that firm-facing actors can create value in use. So, for example, LEGO explicitly sought and harnessed consumer innovation to refine the successful LEGO robotic kit Mindstorms. Bagozzi and Dholakia (2006) investigated the antecedents and purchase consequences of customer participation in brand communities. To disentangle the many variables that play a role in these interactions between community members, brands, and purchases they use Structural Equation Models. The degree by which cocreation actually translates into CLV strongly depends on the trade-off between the costs for a firm to stimulate consumer participation (e.g., financial rewards) and the benefits that firms receive (Hoyer et al. 2010). Such a trade-off needs to be accounted for when incorporating consumer cocreation activities into lifetime value calculations; an issue that becomes even more complex when participation levels are assumed to vary over the customer life cycle (e.g., does the initial motivation to cocreate diminish or does it intensify with relationship duration?). Here, persistence models dealing with time-series data constitute a useful modeling approach for future research.

A further manifestation of customer engagement, which is likely to affect CLV are customer complaints. On the one hand, firms recognize that complaints represent an opportunity to remedy product or service-related problems and to positively influence subsequent customer behavior. There is considerable evidence that dealing effectively with complaints can have a dramatic impact on customers' evaluations of customer experiences (e.g., Bitner, Booms, and Tetreault 1990) as well as enhance their likelihood of repurchase and limit the spread of damaging negative WOM (e.g., Blodgett, Granbois, and Walters 1993; Gilly and Gelb 1982). In this sense, voicing a complaint followed by effective complaint management may positively contribute to CLV. On the other hand, many customers who complain end up feeling more negative about the business because of the way their problems are addressed (Hart, Heskett, and Sasser 1990). In such a case, CLV would be reduced. Hence, modeling this trade-off depending on a firm's complaint management constitutes a fruitful area for future research.

## Models for Customer Retention

Customer retention is the third building block of a customer engagement strategy. Customer retention focuses on preventing customer attrition or churn, that is, the termination of the contractual or noncontractual relationship between the customer and the company. Assessing the churn risk characterizing each customer is also needed for customer valuation. The CLV to a company directly depends on the duration of its relationship with the firm, and so, on the probability of the customer being "alive."

Management of customer retention requires the elaboration of tools that allow managers to assess the risk of each individual to defect. Such tools traditionally identify customers that are the most likely to churn, enabling the allocation of resources across the customer base (Ganesh, Arnold, and Reynolds 2000; Shaffer and Zhang 2002). Managing customer retention also implies an understanding of the factors that trigger customer defection. Several studies have investigated the churn or retention drivers in order to provide companies tools on how to improve the effectiveness of retention programs and hereby prolong the lifetime of customers. For instance, Verhoef (2003) found affective commitment and loyalty programs to reduce churn, while Rust and Zahorik (1993) have found a link between satisfaction and retention, even if this relation might vary across customer segments (Mittal and Kamakura 2001).

In a noncontractual setting, the challenge is to infer whether a customer is still active or not. Most models that have been developed to assess the probability that a customer is still alive are probability models, such as the Pareto/NBD model. In contractual settings (e.g., cellular phones), customer churn is defined as the extinction of the contract between the company and its customer. In this context, the churn problem is traditionally stated as a binary issue, where the aim is to predict whether or not a customer is likely to defect during a pre-given time period. Neslin et al. (2006) provide an overview of the binary models that were used by several academics and practitioners in the context of the churn modeling tournament. Various binary choice models have been used in the past. They include logistic regression analysis, decision trees, and discriminant analysis (see Kamakura et al. 2005 for a review).

Recent developments in the field of machine learning, including neural networks (Hung, Yen, and Wang, 2006), support-vector machine (Coussement and Van den Poel 2008), bagging, and boosting (Lemmens and Croux, 2006), provide a substantial improvement in predictive performance compared to traditional approaches (Risselada, Verhoef, and Bijmolt 2010). In many applications, managers have access to panel data and each customer can be tracked across multiple time points. Substantial improvement in predictive accuracy can also be realized when accounting for the heterogeneity in customer response, for example, using finite mixture models and hierarchical Bayes techniques.

An alternative way to tackle customer churn is to model the duration of customer relationship with the firm. This stream of research uses hazard models to predict the probability of customer defection (Bolton 1998). Gupta and Zeithlaml (2006) classify these models as "lost for good" as they consider customer defection as permanent. A different approach consists of considering customers defection as transient. The "always a share" retention models typically estimate transition probabilities of customers being in a certain state, where customer defection is defined as one of these states. Transition probabilities are estimated using Markov models (Pfeifer and Carraway 2000). Schweidel, Fader, and Bradlow (2008) provide a framework to examine factors underlying service retention in a contractual setting, including duration dependence, promotional effects, subscriber heterogeneity, cross-cohort effects, and calendar-time effects.

*Future Model Directions for Customer Engagement.* Traditional approaches to manage customer retention have focused on predicting which customers are most likely to churn and then target actions to those customers to induce them to stay. However, they have generally mostly ignored the notion of customer engagement to the firm when making such decisions. Numerous tracks for future research open up to researchers interested in linking customer retention decisions with customer engagement.

A promising research area to account for customer engagement in the context of retention is to incorporate the potential interactions across customers in churn prediction models. In the same fashion, as CLV depends on indirect network effects, where customers attract each other, for example, eBay buyers and sellers (Gupta and Mela 2008), the defection of a customer is likely to affect the churn probabilities of other customers of the firm. In other words, one customer terminating his or her contract with the company can lead to cross-customer spillover effect in churn. Such phenomenon is likely to happen in industries such as the telecom where tariffs depend on the provider that customers are calling. Such spillover can also be induced by WOM, as churners are likely to spread out bad publicity among their friends about the company they are leaving. An area for future development might therefore be to incorporate social network information (e.g., who is calling who?) in modeling potential cross-customer spillover effects.

In another direction, retention models would also benefit from accounting for heterogeneity in customer engagement across the customer base. Customers are likely to show different levels of engagement to the firm. However, current modeling approaches ignore the potential heterogeneity across customers in terms of the value they bring to the firm, where customer engagement can be seen as a component of customer value to the firm. Misclassifying a high-value, or a highly engaged, customer is likely to be worse to the firm than misclassifying a low-value, or a low-engagement, customer. A potentially interesting development in the field would be to link a customer engagement and customer value to the firm to his or her churn propensity. One could directly relate the objective function of the method used to the retention program the prediction model is designed for, and hereby design a loss function that would maximize the financial gains of the program (Lemmens and Croux 2010).

Churn modeling could gradually move away from static churn models (e.g., binary choice model) to models that incorporate dynamics into the churn process. In particular, it would be worthwhile to investigate how customer engagement dynamically interacts with customer propensity to churn. While studies have implicitly acknowledged that the churn propensity of customers evolve over time as customers turn for a non-churner state to a churner state (e.g., using hazard modeling), the role of the churn drivers is usually considered stable over a customer lifetime. Future research might consider allowing the churn determinants to vary dynamically along the customer life cycle, for example, using a time-varying coefficient model (Stremersch and Lemmens, 2009) or a dynamic linear model (Ataman, van Heerde, and Mela 2010). Alternatively, one could extend the framework for customer relationship dynamics proposed by Netzer, Lattin, and Srinivasan (2008). Using the proposed hidden Markov model, the probability of churn would depend on the relationship or engagement state a customer experiences at a given time point and the Markovian transitions between states would depend on time-varying covariates.

## Barriers for Implementation in Practice

In this article, we review scientific developments in the area of analytics for customer engagement. These developments provide rich opportunities for application or even exploitation in practice. Yet, a gap between science and practice has been frequently observed (Bemmaor and Franses 2005; Fader and Hardie 2005; Hanssens, Leeflang, and Wittink 2005; Wind and Green 2004). In this section, we discuss important barriers that hinder implementation of customer analytics in practice and potential solutions for solving these barriers.

### *Barrier 1: Data Quality, Size of Databases, and New Types of Data*

The value of customer analytics for supporting management of customer engagement will critically depend on the data available for the modeling tools. Improving the amount and quality of data might very well be more critical than improving the modeling itself. Data quality comprises many aspects of data itself and the services around data, which make it accessible to the data consumer. "Garbage in, garbage out" is a commonly used metaphor with respect to data quality, which also holds for customer analytics.

Data mining became very popular over the last years because it offered a process view on analytics, not focusing exclusively on the statistical modeling of data but also on data sourcing, preprocessing, data transformation, aggregation, and so on. It also has allowed working with thousands of variables to semiautomatically identify potentially good predictors for statistical models. Today, many effects change the way customer analytics is done. Organizations nowadays collect enormous quantities of data. Good examples are banks, telecommunications, insurance companies, energy suppliers, and so on. This represents significant challenges to data management and data quality management. Analytics has to be able to cope with data volumes of any size. Often, main challenge lies in applying a statistical model to a huge sample or even the whole of the population (e.g., the customer base). Urban et al. (2009) describe a Bayesian model for adapting websites to individual customers, which provides an interesting example of such scalable models. To conclude, analytics must be ready to be deployed efficiently on large data sets to provide insights supporting management of customer engagement.

### *Barrier 2: Data Ownership*

Implementation of customer analytics requires that it is clear which department owns the data. Modeling customer engagement requires data, usually from different departments in the firm. Data ownership within an organization defines the members to whom the organization has assigned the responsibility for the asset of customer-related data. This asset is managed by the owner in such a way that she or he exercises all of the organization's rights and interests in the data including the data importance, value, sensitivity, how, and who uses the data.

Increasingly, the perception with respect to the data ownership question has shifted to data privacy aspects. The owner of customer data is, finally, the customer. An organization should be explicitly allowed by their customers to use their data. In addition, the organizations must clearly define to what purposes and by whom customer data may be employed. Therefore, today, instead of using "data ownership," there is a trend to rather naming it "data stewardship."

If we focus on the internal view of an organization on data ownership, that is, looking at who manages customer data, we still often find an over weighted importance of IT-departments. As "data consumers" in marketing usually require a broad view on customer data, especially considering management of customer engagement, in contrast to others who often only require silo-type of transactional data (e.g., billing data to check payment of invoices), a natural place to find data are data warehouses or data marts. These data repositories are usually not in a fully productive operating mode and frequently, given their technical nature, under control of IT. This may make the access to "good quality data" from a marketing data consumer's perspective a cumbersome and time-consuming task, and thereby forms an important barrier to implementing analytical models for customer engagement.

### *Barrier 3: Complexity of the Models*

The methods developed in science are often very complex. Most of the algorithms and statistics are unknown to and too difficult for the manager who should work with the results of the customer analytics. Hanssens, Leeflang, and Wittink (2005) suggest in this respect the development of standardized models. They define a standardized model as a set of one or more relations where the mathematical form and the relevant variables are fixed. A variation consists of the use of subsets of relations as modules. This is attractive if the relevance of modules depends on, say, client factors. In a module-based approach, the structure of each module is fixed. Of course, the estimated equations will often still vary somewhat between applications and over time. For example, predictor variables can be deleted from the relations based on initial empirical results. Standardized models are calibrated with data obtained in a standardized way (audits, panels, and surveys), covering standardized time periods. Outcomes are reported in a standardized format such as tables with predicted own-item sales indices for all possible combinations of display/feature and specific price points or predicted market shares for new products.

### *Barrier 4: Ownership of the Modeling Tools*

Initially, customer analytics was embedded within database marketing, so there was a natural trend to host it in a marketing environment. As the topic developed, it became increasingly dependent on information technology (hardware and software) and databases. This caused a push of customer analytics into IT departments, which certainly had advantages (access to faster computing, more complex software, and more disk space) and drawbacks (loss of customer focus for analytics, too far away from the ‘‘business’’ questions and challenges).

When it comes to the use of customer analytics, there is one component that tends to be overlooked in the literature—the role of the chief information officer (CIO) and his team. Unlike many traditional marketing analytics tools, which rely on third-party databases or primary market research, customer analytics tools depend on the data that lie in the company’s customer databases. Thus, any marketing-driven customer analytics activity requires the buy-in of the CIO and the company’s IT department. Other departments that play a role in submitting data to customer analytics are the accounting department, the department that deals with complaints, the sales force, and the marketing department itself. This is not only for access to the data but also for implementation within the organization. For example, marketing scientists have developed a number of models for identifying cross-sell and up-sell opportunities (Blattberg et al. 2008). Integrating these models, the company’s enterprise systems (e.g., ecommerce website and retail point-of-sale (POS) systems) will be the job of the CIO. While few chief marketing officers (CMOs) have probably read *Competing on Analytics* (Davenport and Harris 2007), it has been getting the attention of CIOs. Marketing scientists interested in the development and implementation of customer analytics tools should be talking to the CIO as much as they talk to the CMO.

Given the role of the CIO when it comes to customer analytics, especially in terms of operational CRM (e.g., Harrahs, see Davenport 2006), it is useful to reflect on the ‘‘supply chain’’ for such tools. If we look at who works in these departments or in the various suppliers on which they rely, we tend to find computer scientists and statisticians, not marketing scientists. It seems that the work on such ‘‘marketing’’ topics as CLV, churn modeling, and experimentation being undertaken by the researchers and consultants coming from these backgrounds (e.g., Kohavi et al. 2008: Rosset et al. 2003) is frequently overlooked by marketing scientists. While some of the models developed by these researchers may lack the ‘‘elegance’’ and ‘‘theoretical foundations’’ desired by marketing scientists, the reality is that they are more likely to make it into the company, given importance of the IT and the CIO.

In general, analytics should be wherever it generates most benefits. Given the central role that marketing plays in customer engagement, we really would argue that the marketing department should be the place where data and tools to analyze customers are managed.

### *Barrier 5: Usability of the Results*

Customer analytics may yield statistical or other results that do not match one to one with the decision to be made by the manager. Often interpretation of the results is too complex for the manager. This implies that the model outcomes should fit the manager’s mental model. In addition, analytics must be able to provide results very fast in order to be ready for intervening at the real-time level to drive decisions. Typical applications are next best product offers generated through analytics on web interfaces in customer-company relationships. Depending on the response time requirements, complex and time-consuming analytical methods may have to be excluded. However, in practice, customer analytics are often too slow to deliver results within desired time frames.

### *Barrier 6: Integration in Company Processes*

Results from analytics need to be taken from the experimental/prototype phase to a production phase, where the models are inserted into current organization processes. This transition involves a number of departments and people with different skill sets, which may reduce the rate at which analytics is deployed within company processes. Ideally, a customer analytic tool should be automated. This is often hard to achieve, which limits its adoption level since costs for model production will be high and deployment within operational processes limited.

Results obtained with the help of analytics need to be generalized in order to be applied to a larger number of cases, situations, and decisions. This will help leverage analytics and increase the benefits-to-cost ratio of analytics. To generalize results from analytics, deep understanding of both analytics and business is required, which leads to involving the corresponding people in this task.

Parallel to integration in the company processes, it is of critical importance that customer analytics has top management support. This will facilitate investment decisions and stimulate adoption of customer analytics throughout the organization. Finally, to support customer engagement management, it is important that the final user of the customer analytics has an important vote in the customer analytics process. The marketing field should play a major role in analytics supporting customer engagement to warrant its added value.

## Conclusion

We have discussed analytical models for customer engagement, which goes beyond models for customer transactions. These models pertain to the subsequent stages of the customer life cycle: customer acquisition, customer development, and customer retention. Important developments regarding data availability (see Section Data for Customer Analytics) allow for more detailed and advanced analysis in each of these stages, which supports management of customer engagement. However, several organizational issues of analytics for customer engagement remain, which constitute barriers for implementing analytics for customer engagement. We anticipate that continuation of the research streams discussed in this article will help to overcome these barriers.

## Authors' Note

This article is a result of the 3rd Thought Leadership Conference on Customer Management, held in Montabaur, Germany, September 2009.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

## Funding

The author(s) received no financial support for the research and/or authorship of this article.

## References

Abe, Makoto (2009), "Counting Your Customers' One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model," *Marketing Science*, 28 (3), 541-553.

Albuquerque, Paulo and Bart Bronnenberg (2009), "Estimating Demand Heterogeneity using Aggregated Data: An Application to the Frozen Pizza Category," *Marketing Science*, 28 (2), 356-372.

Anderson, Eric T. and Duncan I. Simester (2004), "Long-Run Effects of Promotion Depth on New Versus Established Customers: Three Field Studies," *Marketing Science*, 23 (1), 4-20.

Ansari, Asim and Carl F. Mela (2003), "E-Customization," *Journal of Marketing Research*, 40 (2), 131-145.

———, ———, and Scott A. Neslin (2008), "Customer Channel Migration," *Journal of Marketing Research*, 45 (1), 60-76.

Ataman, M. Berk, Harald J. van Heerde, and Carl F. Mela (2010), "The Long-term Effect of Marketing Strategy on Brand Sales, *Journal of Marketing Research* forthcoming.

Baesens, Bart, Stijn Viaene, Dirk Van den Poel, Jan Vanthienen, and Guido Dedene (2002), "Bayesian Neural Network Learning for Repeat Purchase Modelling in Direct Marketing," *European Journal of Operational Research*, 138 (1), 191-211.

Bagozzi, Richard P. and Utpal M. Dholakia (2006), "Open Source Software User Communities: A Study of Participation in Linux User Groups," *Management Science*, 52 (7), 1099-1115.

Balakrishnan, Suhrid and David Madigany (2006), "A One-Pass Sequential Monte Carlo Method for Bayesian Analysis of Massive Datasets," *Bayesian Analysis*, 1 (2), 345-362.

Bass, F. (1969), "A New Product Growth for Model Consumer Durables," *Management Science*, 15 (5), 215-227.

Bauer, Connie L. (1988), "A Direct Mail Customer Purchase Model," *Journal of Direct Marketing*, 2 (3), 16-24.

Bemmaor, Albert C. and Philip Hans Franses (2005), "The Diffusion of Marketing Science in the Practitioners' Community: Opening the Black Box," *Applied Stochastic Models in Business and Industry*, 21 (4/5), 289-301.

Berger, Paul and Thomas Magliozzi (1992), "The Effect of Sample Size and Proportion of Buyers in the Sample on the Performance of List Segmentation Equations Generated by Regression Analysis," *Journal of Direct Marketing*, 6 (1), 13-22.

Bitner, Mary Jo, Bernard H. Booms, and Mary Stanfield Tetreault (1990), "The Service Encounter: Diagnosing Favorable and Unfavorable Incidents," *Journal of Marketing*, 54 (1), 71-84.

Blattberg, Robert C., Byung-Do. Kim, and Scott A. Neslin (2008), *Database Marketing: Analyzing and Managing Customers*. New York: Springer.

Blodgett, Jeffrey G., Donald H. Granbois, and Rockney G. Walters (1993), "The Effects of Perceived Justice on Complainants' Negative Word-of-Mouth Behavior and Repatronage Intentions," *Journal of Retailing*, 69 (4), 399-428.

Bodapati, Anand V. 2008, "Recommendation Systems with Purchase Data," *Journal of Marketing Research*, 45 (1), 77-93.

Bolton, Ruth N. (1998), "A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction," *Marketing Science*, 17 (1), 46-65.

———, Kay N. Lemon, and Peter C. Verhoef (2004), "The Theoretical Underpinnings of Customer Asset Management: A Framework and Propositions for Future Research," *Journal of the Academy of Marketing Science*, 32 (3), 271-292.

Bowman, Douglas and Das Narayandas (2001), "Managing Customer-Initiated Contacts with Manufacturers: The Impact on Share of Category Requirements and Word-of-Mouth Behavior," *Journal of Marketing Research*, 38 (3), 281-97.

——— and ——— (2004), "Linking Customer Management Effort to Customer Profitability in Business Markets," *Journal of Marketing Research*, 41 (4), 433-447.

Bucklin, Randolph E. and Catarina Sismeiro (2009), "Click Here for Internet Insight: Advances in Clickstream Data Analysis in Marketing," *Journal of Interactive Marketing*, 23 (1), 35-48.

Bult, Jan R. and Tom Wansbeek (1995), "Optimal Selection for Direct Mail," *Marketing Science*, 14 (4), 378-394.

Coussement, Kristof and Dirk Van den Poel (2008), "Churn Prediction in Subscription Services: An Application of Support Vector Machines While Comparing Two Parameter-Selection Techniques," *Expert Systems with Applications*, 34 (1), 313-327.

Davenport, Thomas H. (2006), "Competing on Analytics," *Harvard Business Review*, 84 (1), 98-106.

——— and Jeanne G. Harris (2007), *Competing on Analytics: The New Science of Winning*. Boston, MA: Harvard Business School Publishing.

David Shepard Associates (1999), *The New Direct Marketing*, 3rd. New York: Irwin Professional Publishing.

Deshpandé, Rohit (1983), "'Paradigms Lost:' On Theory and Method in Research in Marketing," *Journal of Marketing*, 47 (4), 101-110.

Donkers, Bas, Peter C. Verhoef, and Martijn G. de Jong (2007), "Modeling CLV: A Test of Competing Models in the Insurance Industry," *Quantitative Marketing and Economics*, 5 (2), 163-190.

DuMouchel, William (2002), "Data Squashing: Constructing Summary Data Sets," in *Handbook of Massive Datasets*, James Abello, Panos M. Pardalos, and Muricio G. C. Resende, eds. Dordrecht: Kluwer Academic Publishers, 579-591.

Fader, Peter S. and Bruce G. S. Hardie (2005), "The Value of Simple Models in New Product Forecasting and Customer-Base Analysis," *Applied Stochastic Models in Business and Industry*, 21 (4/5), 461-473.

——— and ——— (2007), "How to Project Customer Retention," *Journal of Interactive Marketing*, 21 (1), 76-90.

——— and ——— (2009), "Probability Models for Customer-Base Analysis," *Journal of Interactive Marketing*, 23 (1), 61-69.

———, ———, and Kinshuk Jerath (2007), "Estimating CLV Using Aggregated Data: The Tuscan Lifestyles Case Revisited," *Journal of Interactive Marketing*, 21 (3), 55-71.

———, ———, and Ka Kok Lee (2005a), "Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, 24 (2), 275-284.

———, ———, and ——— (2005b), "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, 42 (4), 415-430.

———, ———, and Jen Shang (2010), "Customer-Base Analysis in a Discrete-Time Noncontractual Setting," Marketing Science, forthcoming.

Fuchs, Christoph, Emanuela Prandelli, and Martin Schreier (2010), "The Psychological Effects of Empowerment Strategies on Consumers' Product Demand," *Journal of Marketing*, 74 (1), 65-79.

Ganesh, Jaishankar, Mark J. Arnold, and Kristy E. Reynolds (2000), "Understanding the Customer Base of Service Providers: An Examination of the Differences between Switchers and Stayers," *Journal of Marketing*, 65 (3), 65-87.

Gensler, Sonja, Peter S. H. Leeflang, and Bernd Skiera (2010), "Does Online Channel Use Affect Customer-Based Metrics? The Importance to Consider Self-Selection Effects," working paper, Goethe University, Frankfurt; University of Groningen.

Gilly, Mary C. and Betsy D. Gelb (1982), "Post-purchase Consumer Processes and the Complaining Customers," *Journal of Consumer Research*, 9 (3), 323-328.

Goldenberg, Barton J. (2008), *CRM in Real Time: Empowering Customer Relationships*, Medford, NJ: CyberAge Books.

Goldenberg, Jacob, Barak Libai, Sarit Moldovan, and Eitan Muller (2007), "The NPV of Bad News," *International Journal of Research in Marketing*, 24 (3), 186-200.

———, Sangman Han, Donald R. Lehmann, and Jae Weon Hong (2009), "The Role of Hubs in the Adoption Process," *Journal of Marketing*, 73 (2), 1-13.

Gupta, Sunil, Dominique Hanssens, Bruce Hardie, Wiliam Kahn, V. Kumar, Nathaniel Lin, Nalini Ravishanker, and S. Sriram (2006), "Modeling Customer Lifetime Value," *Journal of Service Research*, 9 (2), 139-155.

——— and Carl F. Mela (2008), "What Is a Free Customer Worth?" *Harvard Business Review*, 86 (11), 102-109.

——— and Valarie Zeithaml (2006), "Customer Metrics and Their Impact on Financial Performance," *Marketing Science*, 25 (6), 687-717.

Hansotia, Behram J. and Paul Wang (1997), "Analytical Challenges in Customer Acquisition," *Journal of Direct Marketing*, 11 (2), 7-19.

Hanssens, Dominique, Peter S. H. Leeflang, and Dick R. Wittink (2005), "Marketing Response Models and Marketing Practice," *Applied Stochastic Models in Business and Industry*, 21 (4/5), 423-434.

Hart, Christopher W. L., James L. Heskett, and W. Earl Sasser, Jr. (1990), "The Profitable Art of Service Recovery," *Harvard Business Review*, 68 (4), 148-156.

Hess, P. M., F. R. Kardes, and J. Kim (1991), "Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective," *Journal of Consumer Research*, 17 (4), 454-462.

Hitt, Lorin M. and Frances X. Frei (2002), "Do Better Customers Utilize Electronic Distribution Channels?" *Management Science*, 48 (6), 732-748.

Hoekstra, Janny C. and Peter C. Verhoef (2010), "The Customer Intelligence—Marketing Interface: Its Effect on Firm Performance in Services Organizations," working paper, University of Groningen.

Hoffman, Donna L., Praveen K. Kopalle, and Thomas P. Novak (2010), "The 'Right' Consumers for Better Concepts: Identifying and Using Consumers High in Emergent Nature to Further Develop New Product Concepts," *Journal of Marketing* Research, forthcoming.

Hoyer, Wayne D., Rajesh Chandy, Matilda Dorotic, Manfred Krafft, and Siddarth S. Singh (2010), "Consumer Co-Creation in New Product Development," *Journal of Service Research*, 13 (3), 283-296.

Hung, Shin-Yuan, David C. Yen, and Hsiu-Yu Wang (2006), "Applying Data Mining to Telecom Churn Management," *Expert Systems with Applications*, 31 (3), 515-524.

Kamakura, Wagner A. (2008), "Cross-Selling: Offering the Right Product to the Right Customer at the Right Time," *Journal of Relationship Marketing*, 6 (3/4), 41-58.

———, Carl F. Mela, Asim Ansari, Anand Bodapati, Pete Fader, Raghuram Iyengar, Prasad Naik, Scott Neslin, Baohong Sun, Peter C. Verhoef, Michel Wedel, and Ron Wilcox (2005), "Choice Models and Customer Relationship Management," *Marketing Letters*, 16 (3/4), 279-291.

———, Michel Wedel, Fernando de Rosad, and Jose Afonso Mazzon (2003), "Cross-Selling Through Database Marketing: A Mixed Data Factor Analyzer for Data Augmentation and Prediction," *International Journal of Research in Marketing*, 20 (1), 45-65.

Kaplan, Andreas M. and Michael Haenlein (2010), "Users of the World, Unite! The Challenges and Opportunities of Social Media," *Business Horizons*, 53 (1), 59-68.

Knott, Aaron, Andrew Hayes, and Scott Neslin (2002), "Next-Product-To-Buy Models for Cross-Selling Applications," *Journal of Interactive Marketing*, 16 (3), 59-75.

Kohavi, Ron, Roger Longbotham, Dan Sommerfield, and Randal M. Henne (2008), "Controlled Experiments on the Web: Survey and Practical Guide," *Data Mining and Knowledge Discovery*, 18 (1), 140-181.

Lambert, Diane, José C. Pinheiro, and Don X. Sun (2001), "Estimating Millions of Dynamic Timing Patterns in Real Time," *Journal of the American Statistical Association*, 96 (453), 316-330.

Lemmens, Aurélie and Christophe Croux (2006), "Bagging and Boosting Classification Trees to Predict Churn," *Journal of Marketing Research*, 43 (2), 276-286.

——— and Christophe Croux (2010), "Predicting Customer Churn: Towards a Marketing-Oriented Loss Function," working paper, Erasmus University, Department of Economics, the Netherlands.

Lemon, Kay and Florian Von Wangenheim (2009), "The Reinforcing Effects of Loyalty Program Partnerships and Core Service Usage: A Longitudinal Analysis," *Journal of Service Research*, 11 (4), 357-370.

Lewis, Michael (2005), "Incorporating Strategic Consumer Behavior into Customer Valuation," *Journal of Marketing*, 69 (4), 230-238.

——— (2006), "Customer Acquisition Promotions and Customer Asset Value," *Journal of Marketing Research*, 43 (2), 195-203.

Li, Shibo, Baohong Sun, and Ronald T. Wilcox (2005), "Cross-Selling Sequentially Ordered Products: An Application to Consumer Banking Services," *Journal of Marketing Research*, 42 (2), 233-239.

Mahajan, Vijay, Eitan Muller, and Yoram Wind (2000), *New-Product Diffusion Models*. Boston: Kluwer.

Malthouse, Edward C. (2001), "Assessing the Performance of Direct Marketing Scoring Models," *Journal of Interactive Marketing*, 15 (1), 49-62.

——— and Robert C. Blattberg (2005), "Can We Predict Customer Lifetime Value?" *Journal of Interactive Marketing*, 19 (1), 2-16.

McDermott, James P., G. Jogesh Babu, John C. Liechty, and Dennis K. J. Lin (2007), "Data Skeletons: Simultaneous Estimation of Multiple Quantiles for Massive Streaming Datasets with Applications to Density Estimation," *Statistical Computing*, 17 (4), 311-321.

Mittal, Vittas and Wagner Kamakura (2001), "Satisfaction, Repurchase Intent and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics," *Journal of Marketing Research*, 38 (1), 131-142.

Mobasher, Bamshad, Robert Cooley, and Jaideep Srivastava (2000), "Automatic Personalization Based on Web Usage Mining," *Communications of the ACM*, 43 (8), 142-151.

———, Honghua Dai, Tao Kuo, and Miki Nakagawa (2001), "Effective Personalization Based on Association Rule Discover from Web Usage Data," *Proceedings of WIDM 2001*, *3rd ACM Workshop on Web Information and Data Management*. Atlanta, GA, USA, 9-15.

Moe, Wendy W. and Peter S. Fader (2004), "Dynamic Conversion Behavior at E-Commerce Sites," *Management Science*, 50 (3), 326-335.

Montgomery, Alan L., Shibo Li, Kannan Srinivasan, and John C. Liechty (2004), "Modeling Online Browsing and Path Analysis Using Clickstream Data," *Marketing Science*, 23 (4), 579-595.

——— and Michael D. Smith (2009), "Prospects for Personalization on the Internet," *Journal of Interactive Marketing*, 23 (2), 130-137.

Nasraoui, Olfa and Christopher Petenes (2003), "Combining Web Usage Mining and Fuzzy Inference for Website Personalization," *Proceedings of WebKDD 2003 – KDD Workshop on Web Mining as a Premise to Effective and Intelligent Web Applications*. Washington DC, 37, August 2003.

Neslin, Scott. A., Sunil Gupta, Wagner Kamakura, Lu Junxiang, and Charlotte H. Mason (2006), "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models," *Journal of Marketing Research*, 43 (2), 204-211.

Netzer, Oded, James M. Lattin, and V. Srinivasan (2008), "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science*, 27 (2), 185-204.

Pfeifer, Phillip and Robert Carraway (2000), "Modeling Customer Relationships as Markov Chains," *Journal of Interactive Marketing*, 14 (2), 43-55.

Reinartz, Werner, Manfred Krafft, and Wayne D. Hoyer (2004), "The Customer Relationship Management Process:Its Measurement and Impact on Performance," *Journal of Marketing Research*, 41 (3), 293-305.

———, Jacquelyn S. Thomas, and Ganaél Bascoul (2008), "Investigating Cross-Buying and Customer Loyalty," *Journal of Interactive Marketing*, 22 (1), 5-20.

———, ———, and V. Kumar (2005), "Balancing Acquisition and Retention Resources to Maximize Customer Profitability," *Journal of Marketing*, 69 (1), 63-79.

——— and Rajkumar Venkatesan (2008), "Models for Customer Relationship Management (CRM)" in *Handbook of Marketing Decision Models*, Berend Wierenga, ed. New York: Springer Science and Business Media, 291-326.

Risselada, Hans, Peter C. Verhoef, and Tammo H. A. Bijmolt (2010), "Staying Power of Churn Prediction Models," *Journal of Interactive Marketing*, forthcoming.

Rosset, Saharon, Einat Neumann, Uri Eick, and Nurit Vatnik (2003), "Customer Lifetime Value Models for Decision Support," *Data Mining and Knowledge Discovery*, **7** (July), 321-339.

Rossi, Peter E., Greg M. Allenby, and Rob McCulloch (2005), *Bayesian Statistics and Marketing*. Hoboken, NJ: John Wiley.

Rust, Roland T. and Anthony J. Zahorik (1993), "Customer Satisfaction, Customer Retention and Market Share," *Journal of Retailing*, 69 (2), 193-215.

Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), "Counting Your Customers: Who are They and What Will They do Next?" *Management Science*, 33 (1), 1-24.

Schweidel, David A., Peter Fader, and Eric Bradlow (2008), "Understanding Service Retention Within and Across Cohorts using Limited Information," *Journal of Marketing*, 72 (1), 82-94.

Shaffer, Greg and John Z. Zhang (2002), "Competitive One-to-One Promotions," *Management Science*, 48 (9), 1143-1160.

Sismeiro, Catarina and Randolph E. Bucklin (2004), "Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach," *Journal of Marketing Research*, 41 (3), 306-323.

Stremersch, Stefan and Aurélie Lemmens (2009), "Sales Growth of New Pharmaceuticals across the Globe: The Role of Regulatory Regimes," *Marketing Science*, 28 (4), 690-708.

Thomas, Jacquelyn S., Robert C. Blattberg, and Edward J. Fox (2004), "Recapturing Lost Customers," *Journal of Marketing Research*, 41 (1), 31-45.

Ting, I-Hsien, Chris Kimble, and Daniel Kudenko (2005), "UBB Mining: Finding Unexpected Browsing Behaviour in Clickstream Data to improve a Web Site's Design," *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence* Compiégne University of Technology, France, 179-185..

Tirenni, Giuliano, Abderrahim Labbi, Cesar Berrospi, Andre Elisseeff, Timir Bhose, Kari Pauro, and Seppo Poyhonen 2007, "Customer Equity and Lifetime Management (CELM) Finnair Case Study," *Marketing Science*, 26 (4), 553-565.

Treacy, M. and Fred Wiersema (1993), "Customer Intimacy and Other Value Disciplines," *Harvard Business Review*, 71 (1), 84-93.

Trusov, M., Randolph Bucklin, and Koen Pauwels (2009), "Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing*, 73 (5), 90-102.

Urban, Glen L, John R. Hauser, Guilherme Liberali, Michael Braun, and Fareena Sultan (2009), "Morph the Web to Build Empathy, Trust, and Sales," *MIT Sloan Management Review*, 50 (4), 53-61.

Van Doorn, Jenny, Katherine Lemon, Vikas Mittal, Stephan Nass, Doreen Pick, Peter Pimer, and Peter Verhoef (2010), "Customer Engagement Behavior: Theoretical Foundations and Research Directions," *Journal of Service Research*, 13 (3), 253-266.

Van Eck, Peter, Wander Jager, and Peter S. H. Leeflang (2010), "Opinion Leaders' Role in Innovation Diffusion: A Simulation Study," working paper, University of Groningen, the Netherlands.

Venkatesan, Rajkumar and V. Kumar (2004), "A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy," *Journal of Marketing*, 68 (4), 106-125.

———, ———, and T. Bohling (2007), "Optimal CRM Using Bayesian Decision Theory; An Application to Customer Selection," *Journal of Marketing Research*, 44 (4), 579-594.

———, ———, and N. Ravishanker (2007), "Multichannel Shopping: Causes and Consequences," *Journal of Marketing Research*, 72 (2), 114-132.

Verhoef, Peter C. (2003), "Understanding the Effect of Customer Relationship Management Efforts on Customer Retention and Customer Share Development," *Journal of Marketing*, 67 (4), 30-45.

——— and Bas Donkers (2005), "The Effect of Acquisition Channels on Customer Loyalty and Cross-Buying," *Journal of Interactive Marketing*, 19 (2), 31-43.

———, Jenny van Doorn, and Matilda Dorotic (2007), "Customer Value Management: An Overview," *Marketing: Journal of Research and Management*, 2, 51-68.

Villanueva, Julian, Shijin Yoo, and Dominique M. Hanssens (2008), "The Impact of Marketing-Induced Versus Word-of Mouth Customer Acquisition on Customer Equity Growth," *Journal of Marketing Research*, 45 (1), 48-59.

Vroomen, Björn, Bas Donkers, Peter C. Verhoef, and Philip H. Franses (2005), "Selecting Profitable Customers for Complex Services on the Internet," *Journal of Service Research*, 8 (1), 37-47.

Wangenheim, Florian von and Tomás Bayón (2007), "The Chain from Customer Satisfaction via Word-of-mouth Referrals to New Customer Acquisition," *Journal of the Academy of Marketing Science*, 35 (2), 233-249.

Wind, Yoram and Paul E. Green (2004), "Reflections and Conclusions: The Link Between Advances in Marketing Research and Practice," in *Market Research and Modeling: Progress and Prospects*, Jerry Wind and Paul E. Green, eds. Boston, MA: Kluwer Academic Publishers, 301-317.

Yang, Sha and Greg M. Allenby (2003), "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research*, 40 (3), 282-294.

## Bios

**Tammo H.A. Bijmolt** is Professor of Marketing Research at the Department of Marketing, Faculty of Economics and Business Administration, University of Groningen, the Netherlands. In addition, he is director of the research school SOM of that faculty. His research interests include conceptual and methodological issues such as loyalty programs, customer relationship management, retailing, and meta-analysis. His publications have appeared in among others Journal of Marketing, Journal of Marketing Research, Journal of Consumer Research, International Journal of Research in Marketing, Psychometrika, and the Journal of the Royal Statistical Society (A). He serves on the editorial boards of International Journal of Research in Marketing and the International Journal of Electronic Commerce.

**Peter S. H. Leeflang** is the Frank M. Bass Professor of Marketing at the University of Groningen and member of the Royal Netherlands Academy of Arts and Sciences. Since March 2010 he holds the BAT-chair in Marketing at the LUISS Guido Carli University in Rome. He has authored or co-authored 25 books and published extensively in, inter alia, The Journal of Marketing, The Journal of Marketing Research (JMR), The International Journal of Research in Marketing (IJRM), Management Science, Marketing Science (MsC), Quantitative Marketing and Economics, and the Journal of Econometrics and he serves on the editorial boards of JMR, MsC and IJRM. He is also Guest Professor at the Johann Wolfgang Goethe Universitðt at Frankfurt am Main and at the Hochschule St. Gallen in Switzerland.

**Frank Block** is CEO and founder of FinScore SA, and is an experienced CRM and Data Mining practitioner. He has successfully managed several large, strategic projects and knows how to Data Mining in business practice such as to maximally leverage its benefits. The range of experience includes designing and building large data warehouses, implementing business intelligence solutions, designing and introducing campaign management systems, leading information quality task force groups, developing and applying data mining methodologies in practice.

**Maik Eisenbeiss** is the OBI Assistant Professor of Marketing and Retailing at the Department of Retailing and Customer Management at the University of Cologne, Germany. He received his PhD from the University of Muenster, Germany. His research interests include customer management, retailing, and marketing channels.

**Bruce G. S. Hardie** is Professor of Marketing at London Business School. He holds B.Com and M.Com degrees from the University of Auckland (New Zealand), and MA and PhD degrees from the University of Pennsylvania. His primary research interest lies in the development of data-based models to support marketing analysts and decision makers, with a particular interest in models that are easy to implement. Most of his current projects focus on the development of probability models for customer-base analysis. Bruce's research has appeared in various marketing, operations research and statistics journals, and he currently serves on the editorial boards of the International Journal of Research in Marketing, the Journal of Interactive Marketing, and Marketing Science.

**Aurélie Lemmens** (Ph.D., K.U.Leuven) is an Associate Professor at Erasmus University Rotterdam (The Netherlands). Her research interests are in international marketing and econometric modeling. Her work has appeared in leading scholarly journals, such as Marketing

Science, Journal of Marketing Research, and International Journal of Research in Marketing. She also received financial support from the European Union in the framework of the Marie Curie Individual Fellowship (project 41651) and from the Dutch Science Foundation (NWO) in the framework of a VENI grant (project 451-09-005).

**Peter Saffert** is Ph.D. student at the Department of Retailing and Customer Management at the University of Cologne, Germany. He holds a diploma (masters equivalent) in Econo-mathematics from the University of Ulm, Germany. His research interests include market-response modeling and customer relationship management.