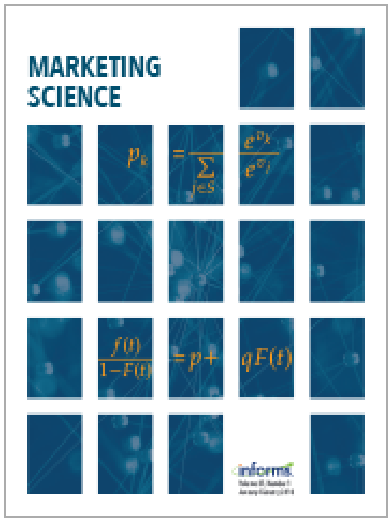


Marketing Science

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

Path Data in Marketing: An Integrative Framework and Prospectus for Model Building

Sam K. Hui, Peter S. Fader, Eric T. Bradlow,

MARKETING SCIENCE
Vol. 28, No. 2, March–April 2009, pp. 320–335
ISSN 0732-2399 | EISSN 1526-548X | 09 | 2802 | 0320

DOI 10.1287/mksc.1080.0400

Path Data in Marketing: An Integrative Framework and Prospectus for Model Building

Sam K. Hui
Stern School of Business, New York University, New York, New York 10012, khui@stern.nyu.edu

Peter S. Fader, Eric T. Bradlow
The Wharton School of the University of Pennsylvania, Philadelphia, Pennsylvania 19104
{faderp@wharton.upenn.edu, ebradlow@wharton.upenn.edu}

Many data sets, from different and seemingly unrelated marketing domains, all involve *paths*—records of consumers' movements in a spatial configuration. Path data contain valuable information for marketing researchers because they describe how consumers interact with their environment and make dynamic choices. As data collection technologies improve and researchers continue to ask deeper questions about consumers' motivations and behaviors, *path data* sets will become more common and will play a more central role in marketing research.

To guide future research in this area, we review the previous literature, propose a formal definition of a path (in a marketing context), and derive a unifying framework that allows us to classify different kinds of paths. We identify and discuss two primary dimensions (characteristics of the spatial configuration and the agent) as well as six underlying subdimensions. Based on this framework, we cover a range of important operational issues that should be taken into account as researchers begin to build formal models of path-related phenomena. We close with a brief look into the future of path-based models, and a call for researchers to address some of these emerging issues.

Key words: path data; path models; integrative review; grocery shopping; eye tracking; Web browsing; clickstream; information acceleration
History: Received: February 9, 2007; accepted: February 27, 2008; processed by Alan Montgomery. Published online in *Articles in Advance* October 9, 2008.

1. Introduction

Consider the following domains of data collection and active research in marketing:

(1) *Grocery shopping*: A grocery store installs radio-frequency identification (RFID) tags underneath shopping carts to track consumers' in-store movements.

(2) *Eye tracking*: Researchers ask subjects to view print advertisements and capture their eye movements using infrared corneal reflection technology.

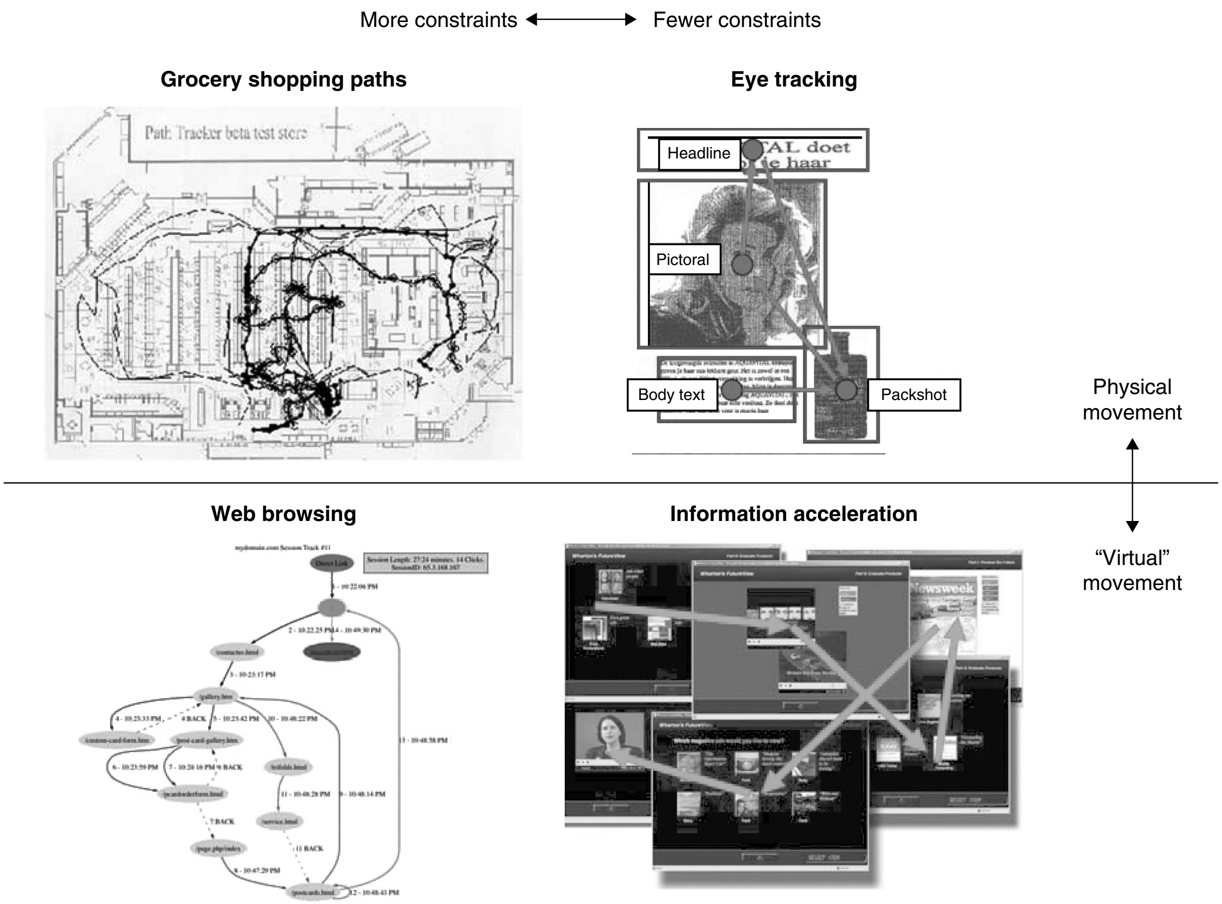
(3) *Web browsing*: Consumers' Web browsing patterns are tracked by recording the sequence of Web pages visited in each session.

(4) *Information acceleration* (Urban et al. 1997): Researchers immerse subjects in a multimedia environment (often with an underlying experimental design) to understand how they collect information and make decisions about a radically new product or service.

At first thought, these domains may not seem to have much in common: they involve different settings and describe different kinds of behaviors. For instance, while grocery shopping and eye tracking involve actual physical movements, Web browsing and information acceleration (IA) occur only in virtual environments. In addition, consumers' underlying goals will generally be quite different in each of these areas; in some cases, the consumer's aim may be to purchase products, while in other cases the main goal is to gather information. Despite these surface differences, a cursory look at Figure 1 suggests that all of these domains share a common theme: they all involve the collection and analysis of different kinds of path data.

Path data (which will be defined formally in §2) create a record of a person's movement in a spatial configuration. They show how a consumer interacts with his environment to achieve his goal(s). In this paper, we argue that there is a need for a closer focus on, and integration of, path-related research in marketing for three reasons. First, the study of paths, in conjunction with other data sources (e.g., surveys, transaction data), may lead to a deeper understanding of consumer behavior; e.g., consumers' paths may be

Figure 1 Four Examples of Path Data



informative about their decision processes and goal orientations. For instance, Montgomery et al. (2004) demonstrated that by taking into account a consumer's page-to-page *browsing* path, online *purchase* behavior can be predicted more accurately (i.e., paths help predict choices). Second, understanding the paths that consumers follow is, in some cases, the primary dependent variable of interest by itself (Bradlow et al. 2005). For example, understanding store traffic patterns may help retailers optimize store layout (Vrechopoulos et al. 2004); similarly, studying eye movements may lead to insights on how consumers shift their visual attention and thus inform advertisers on how to design ads to maximize their impact (Fox et al. 1998). Finally, as we demonstrate in §3, path data are emerging in many areas of marketing. As data collection technologies continue to improve, we will likely see an explosive growth in path-related research. Thus, it is important to understand the underlying dimensions and associated modeling challenges for path data, and to take full advantage of these new data opportunities.

Although marketing researchers have (implicitly) collected and studied path data, the path aspect of the data is usually underexploited, and the commonalities across domains are rarely recognized and appreciated. What is missing, we believe, is a general framework that organizes and unifies path-related marketing contexts. Such a framework will serve three important research purposes. First, it allows us to categorize different path data to understand the similarities and differences among them; this will be helpful for the researcher who is looking for analogous models and managerial settings. Second, with a unified framework, modelers can identify a set of common issues in the analysis and modeling of paths, so that researchers can tackle problems that arise across different areas simultaneously. Third, a formal framework will help researchers borrow the tools developed in other disciplines (e.g., models of animal or pedestrian movement) to analyze marketing paths.

The central goal of our paper is to provide such a framework. In the next section, we define the kinds of paths that make up the focus of this paper and provide a formal definition of a path that is grounded in real analysis (Rudin 1976) and graph theory (Bollobas 1979). We then describe our general framework in detail, focusing on two key components: the nature of the spatial configuration and the characteristics of the agent who follows the path of interest. In §3, we review the different types of path data that arise in four active areas of marketing research: retail/service environments, advertising studies, e-commerce, and experimental research. We also briefly review the literature outside of marketing to help describe the broader context and framework that we develop. Based on our framework and literature review, we identify modeling issues that will commonly arise for researchers who are attempting to understand or predict path behavior. We conclude by offering some observations and recommendations for future research.

2. A Framework for Path Data

In this paper, we focus on paths that arise in a marketing context.

Definition. A path is a conscious (Blackmore 2003) agent's movement in a physical or simulated (Kalawsky 1993) environment that is observable.[1]

A path can be denoted by a three-tuple $P = \{S, A, X_A(t)\}$. S denotes an observable, physical or simulated (e.g., Web, IA) environment; it represents a spatial configuration that contains all possible locations that can be realized.[2] The existence of an observable,[3] physical/simulated environment is an important feature of path data that differentiates them from other sequence data (e.g., brand choice data) where an explicit spatial environment is not available. S may be continuous, i.e., a fixed subset of $\Re^r$ that contains an r-dimensional rectangle of positive volume (Banerjee et al. 2004), or it can be discrete, in which case it would be defined by a mathematical graph.[4] We discuss this differentiation in detail in §2.1.2.

The second component of the three-tuple, A, denotes a conscious agent who is making the movements. Because this is our focus, other path-related research areas such as robot paths (e.g., Thrun et al. 2005) or hurricane paths (e.g., Bril 1995) are not considered in this paper. While both of the above areas contain many useful insights, we believe that our focus on conscious agents is reasonable given that our goal is to analyze paths made by consumers in various marketing contexts.

Once S and A are specified, X_A denotes the movement of the agent. Specifically, $X_A(t)$ represents the position of the agent A at time t within the spatial configuration. For P to be considered a path, $X_A(t)$ must vary continuously over time with respect to the spatial configuration. If S is continuous, the continuity of $X_A(t)$ can be defined as the conventional definition of continuity in mathematical analysis (Rudin 1976), given by

$$\lim_{\Delta t \to 0} X_A(t + \Delta t) = X_A(t). \qquad (1)$$

On the other hand, if S is discrete (a graph), $X_A(t)$ must vary in a way that is consistent with the structure of the graph; i.e., subsequent moves must be between nodes connected by an edge.[5] This important definitional element marks the key difference between path data and other kinds of space-time data. More specifically, spatio-temporal data sets, such as the product diffusion data in Bell and Song (2007), do not belong to path data under our definition. In product diffusion data, the data are not continuous: two successive adopters can come from very distant locations.[6] Thus, although those data also have both spatial and temporal elements, they do not conform to the continuity restriction and are thus not considered to be path data.

In the following discussion, we outline the two key components of a path model (the spatial configuration S and the role of the agent A) and discuss the specific subdimensions on which data considerations (and thus model specifications) may differ.

2.1. Spatial Configuration (S)

Spatial configuration defines the space in which movements are made and specifies the set of allowable movements given an agent's current location. We characterize a spatial configuration along three separate subdimensions: (1) physical/nonphysical, (2) continuous/discrete, and (3) the presence/degree of constraints. Figure 2 illustrates spatial configurations in six different settings.

2.1.1. Physical/Nonphysical. A spatial configuration may or may not correspond to an actual physical space. As examples of physical spaces, in a model of bird movement, the space is the three-dimensional

[1] Throughout this paper, we focus on an individual-level analysis of paths. Hence, macro level modeling of paths, e.g., fluid dynamic models (Henderson 1974) or cellular automata descriptions (Chopard et al. 1996) are not considered.

[2] Note that our definition of path is on the underlying *process*, not on its *measurement*. The particular measurement of a path is subject to measurement error and is limited by the sampling frequency.

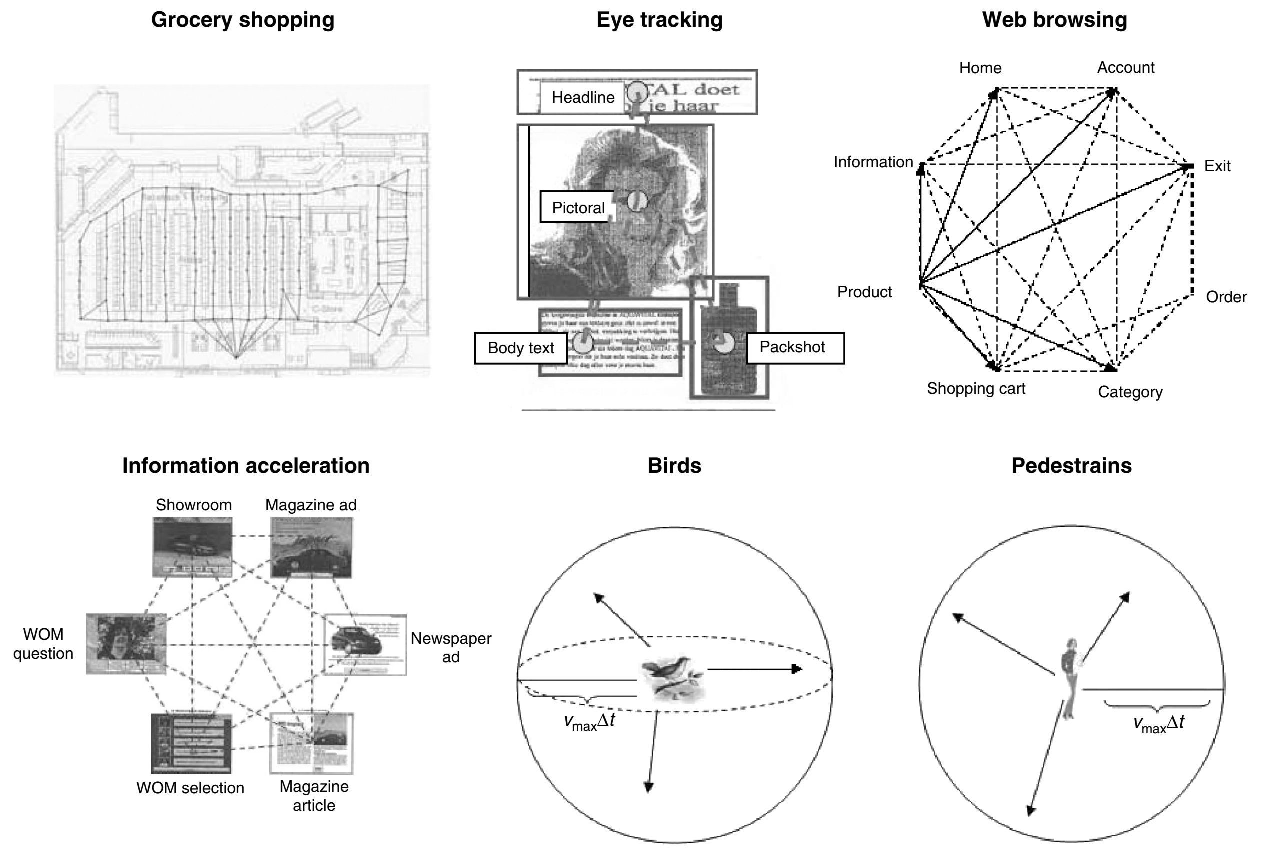
[3] Paths that occur in latent space, e.g., a consumer's change of lifecycle state in a hidden Markov model (Du and Kamakura 2006), are not considered.

[4] A graph is a mathematic object that consists of nodes and edges. Nodes correspond roughly to different locations; the existence of an edge between two nodes indicates that it is possible to move from one node to the other in one step. For details, refer to Bollobas (1979).

[5] In graph theory literature, this defines a walk. We call it a path because it is more consistent with the terminology used in this exposition and in the marketing literature.

[6] One could argue that if there is a fixed and known social structure that connects different consumers, then the continuity restriction may be met. In that case, however, the data correspond to the movement of multiple agents, instead of an individual agent, which is the focus of this paper.

Figure 2 Space and Allowable Movements



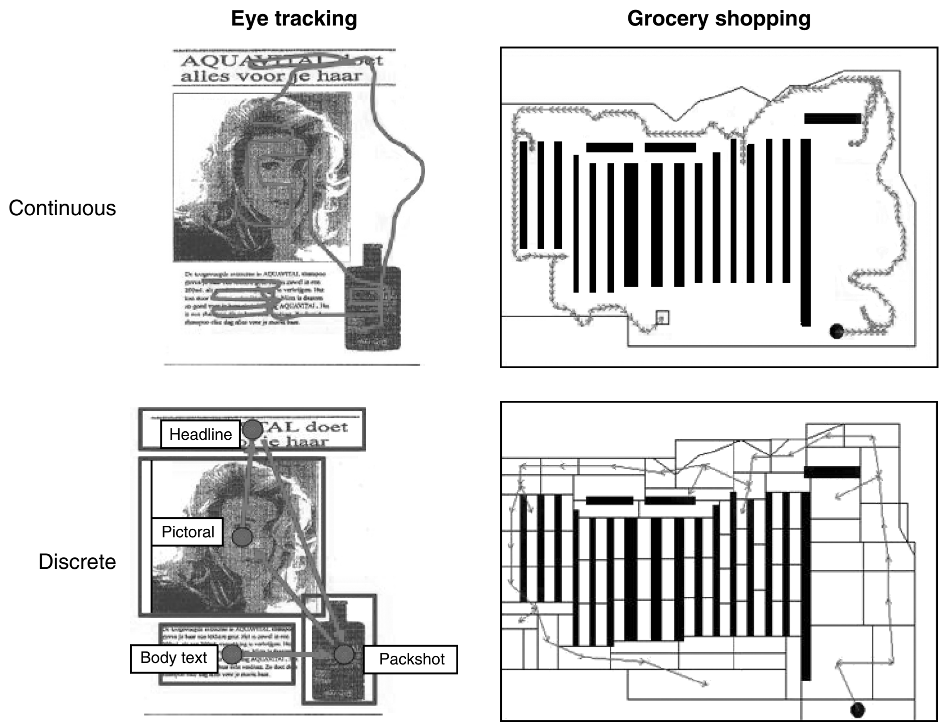
sky; for eye-tracking data, it is the two-dimensional physical image that the subject is viewing. Pedestrian movements also use a space that is generally limited to two dimensions; likewise for a grocery shopper, although the space of movements also includes a number of fixed impediments (aisles, display cases, etc.), which we will discuss shortly.

Some marketing path data, however, do not take place in a physical environment, but instead in a computer-simulated environment (Kalawsky 1993, Hoffman and Novak 1996); examples include Web browsing and IA sessions. In these cases, we often need to generalize the notion of space to incorporate nonphysical spatial configurations. This requires a careful mathematical depiction. For instance, Montgomery et al. (2004) built a discrete spatial configuration for Web browsing: a graph consisting of eight nodes, with each node representing one of the eight online Web categories, as shown in the top right panel of Figure 2. Using this representation, each category can be viewed as a location on a (hypothetical) map defined by Web pages, and therefore each consumer's Web browsing sequence can be treated as a path on this map. This convenient representation is frequently used to model the structure of the World Wide Web (Broder et al. 2000, Eirinaki et al. 2005). Similarly, we can construct a graph, as shown on the bottom left panel of Figure 2, to represent the spatial configuration of data collected from an IA session. Each location on the graph represents a source of information available to the consumer. Thus, the consumer's sequence of information viewing can be seen as a path on this graph. To the best of our knowledge, this characterization has never been used to analyze and describe IA data.

2.1.2. Continuous/Discrete. A spatial configuration can be either continuous or discrete. In models of birds and pedestrians, the spatial configurations are naturally continuous. In contrast, the spatial configurations for Web browsing data or IA are inherently discrete because, as above, the space is constructed as a finite-node graph.

In some applications, however, the modeler can decide whether a continuous or a discrete representation is more suitable. For example, the space for eye tracking data can be modeled as continuous if we use the entire advertisement as the space, and model the movement of the visual focus between two pairs of (x, y) coordinates (see Figure 3, top left panel). Alternatively, it can be modeled as discrete if we characterize the space as the different elements (e.g., picture, brand logo, text) in the advertisement, and model the transition of a consumer's visual focus among them

Figure 3 Discrete vs. Continuous Specification of Eye Tracking and Grocery Shopping



(Pieters and Wedel 2004), as depicted in the bottom left panel of Figure 3. Similarly, we can define the space of grocery shopping as continuous by defining it in terms of (x, y) coordinates in the store, or we can define it as discrete by modeling shopper's movement between zones or departments (see top and bottom right panels, Figure 3). This *discretization*, approximating a continuous object by a mesh of discrete points, is similar to the use of the finite element method in engineering (Zienkiewicz et al. 2005) and is related to aggregation issues or (wombling) in spatial statistics (Banerjee et al. 2004).

On the one hand, discretization may reduce the complexity of the data and thus facilitate computation; on the other hand, it is often not clear what level of aggregation is most suitable for a specific problem. We return to this issue and discuss the relative advantages of a discrete versus continuous spatial specification in more detail later when we outline model-building issues for marketing researchers.

2.1.3. Presence/Degree of Constraints. Constraints may be present in a spatial configuration where they serve as restrictions on the set of allowable movements that an agent, A, may take. In a physical setting, constraints are usually characterized by physical impediments. For example, aisles and walls in a grocery store restrict the possible direction of movement for grocery shoppers. Constraints may be less severe in other settings. For instance, in a model of pedestrian movement, the only constraints that a pedestrian faces is that he or she cannot walk into physical objects. Other than that, an allowable movement is any location within the circle with the person's current location as the center and a radius equal to $v_{\max}\Delta t$, where $v_{\max}$ is the person's maximum walking speed (Helbing et al. 1997). Likewise, physical constraints are generally absent in eye tracking studies.

Note that in physical spatial configurations there is a natural relationship between the degree of constraints and whether the space should be treated as continuous or discrete. In general, it is more appropriate to model space as discrete when movements are highly constrained. That is, movements are viewed as a sequence of choices among a finite set of alternatives, rather than a choice among infinitely many directions. As an example, in the pedestrian model, because of the lack of constraints, it may be more appropriate to model movements as occurring continuously in space; in contrast, for grocery shopping, it may be more appropriate to model the paths as transitions between different store zones (Hui et al. 2007).

Constraints may also exist in nonphysical settings. If we represent a nonphysical setting as a graph, the presence of constraints between two points implies the absence of edges between the respective nodes that represent them. For instance, in Montgomery et al. (2004), at each step the user is located at the node that represents his current category, and he is allowed to move to any node that is connected. As can be seen on the top right panel of Figure 2, the product node and the order node are not directly connected. This implies the existence of a constraint—a person cannot directly reach the *confirm order* node from the product node in one step because he has to go through the *shopping cart* node in between. In IA sessions, constraints are largely absent and the agent is relatively free to move to any location. Note, however, that sessions are often set up in such a way that not all possible movements are available in one step.

2.2. The Agent (A)

Having defined a spatial configuration, researchers need to characterize the conscious agent who is making the movements. For most path models, the definition of an agent is usually straightforward. For example, the agent for a grocery shopping trip is the shopper; for a Web browsing model, the agent is the user surfing the Internet. Sometimes the agent is a well-defined group, e.g., a family traveling in a car. Complications can arise if a multiunit agent takes turns, e.g., a parent who lets a child push the grocery cart, or if the data collection process loses track of specific entities (which can conceivably occur with birds, for instance). One of the clear advantages of experimental settings (such as IA and eye tracking) is the high degree of control that the researcher maintains in this regard.

Beyond these definitional aspects, there are three key subdimensions that should be considered for each agent: (1) social interaction, (2) goal-directedness, and (3) forward-looking tendencies, all active research topics in marketing today.

2.2.1. Social Interaction. When the agent co-occupies the same spatial configuration with other agents, social effects may be present. That is, an agent may be affected by the other agents' actions when deciding his next movement. For instance, although each bird in a flock is autonomous, it tends to match the velocity of its nearby flockmates to avoid collision, while staying close to the center of the flock (Reynolds 1987). Similarly, researchers found that pedestrians have strong tendencies to preserve interpersonal spaces to avoid running into each other (Helbing et al. 1997).

Social interaction between shoppers is especially relevant in the physical retail environment. For instance, researchers found that grocery shoppers typically avoid walking into crowded areas (Harrell et al. 1980) and have demonstrated that even the "mere social presence" of others will affect shopping behavior (Argo et al. 2005). In particular, consumers may reduce their shopping time (Harrell et al. 1980) or even refrain from purchasing a product if they feel that their personal space is "invaded" (Underhill 2000). On the other hand, consumers may exhibit a certain herding behavior by moving toward areas where other shoppers are heading, presumably because shoppers may infer the attractiveness of a store location based on other shoppers' actions (e.g., Banerjee 1992). For instance, Becker (1991) documented that when given the choice between two almost identical restaurants, people usually choose the one with a longer line despite the longer wait. This leads to an interesting dynamic: while retailers should carefully design their store to avoid overcrowding conditions, they may also want to make sure that certain areas do not receive too little traffic. Thus, social interactions between shoppers may be an important issue to be considered when designing a retail environment.

In some cases, however, social effects may be insignificant. Web browsing, participation in an IA session, and reading an advertisement during an eye tracking study are all individual activities that, by design, involve little interaction among agents. Thus, Montgomery et al. (2004) did not choose to consider social effects in their model of Web browsing data. Similarly, researchers can ignore any social effects when modeling path data from eye tracking and IA sessions. At the same time, however, one may wonder whether movements in both of these cases would be influenced by social norms (of others), and/or expectations by the participant as to what he or she is supposed to look at (social cues).

Information acceleration commonly attempts to bring in social cues through personal testimonials, which can influence the subsequent paths chosen by the subject. This is an intriguing aspect of technology, particularly when these testimonials (and the characteristics of the person(s) delivering them) are manipulated experimentally. But while the cues may affect subjects' movement, there is no provision for *direct* social influence in current applications of IA. This may be possible, however, by including statements such as "80% of previous users clicked on area Y after visiting this page." Participants may become more inclined to view area Y after viewing this suggestive statement.

2.2.2. Goal-Directedness. An agent may be goal-directed (Lee and Ariely 2006) and begin the path process with specific goal(s) in mind (e.g., to buy groceries for dinner, to find a textbook on Amazon.com), or at the other extreme, his trip can be a purely hedonic browsing experience. Agents with varying degrees of goal-directedness may exhibit very different paths. For instance, in her study of online shoppers, Moe (2003) identified four types of Internet browsers with different goals, directed buying, search and deliberation, knowledge building, and hedonic browsing, and found that distinct types of shoppers exhibit very different page-to-page browsing patterns (paths). This distinction can be extended to other areas that involve paths. For example, a grocery shopper with a list probably follows a different movement pattern than someone who is primarily window shopping. Likewise, a person reading a magazine to find product information may exhibit a different eye-movement pattern than someone who is leisurely reading an article. In online settings, Web users also exhibit different degrees of goal-directed behavior. For instance, Montgomery et al. (2004) allowed each subject to be in either a browsing or a deliberative latent state. In each state, the subject exhibited different search behavior. Thus, to accurately specify a model of movements, researchers must carefully define what goal-directedness means in each context and capture differences across agents on this dimension. This is an example of heterogeneity among agents, an issue we will return to in §4.

Note, however, that path data alone do not always give reliable indications of the mood, goals, and intentions of the agents. Thus, whenever possible, we recommend that other methods (e.g., surveys) be used in conjunction with path data to capture that information.

2.2.3. Forward-Looking Behavior. A goal-directed agent can be forward-looking (e.g., Song and Chintagunta 2003, Sun et al. 2003). That is, along his path, an agent may plan ahead for subsequent actions. Researchers have considered various degrees of planning-ahead behavior in many path settings. For instance, Hui et al. (2008) studied how a grocery shopper plans ahead the order in which he gathers products on his shopping list. At a more macro level, researchers in the trip chaining literature studied how consumers plan ahead to chain together different errands in different locations (e.g., Adler and Ben-Akiva 1979).

Forward-looking behaviors are manifested not only in retail settings, but also in situations that involve information search. A Web user may have a specific search plan/query in mind when trying to satisfy information needs (Robertson 1977). For example, a person seeking to purchase a digital camera may first plan to collect information on the types of cameras available, then on specific features (e.g., resolution, battery life), then decide on the exact model, and finally conduct a price search on that model. Even in a nonpurchase setting such as IA, a person may plan ahead on how he wants to evaluate the products/services of interest. For instance, he may plan to first pay attention to the objective features of the products, before taking into account the more subjective information such as product reviews and word-of-mouth (WOM) information.

While the extent of plan-ahead behavior is an important characteristic of the agent, different agents may exhibit different degrees of forward-looking tendencies. To realistically capture consumer behavior, path modelers need to specify and incorporate different degrees of forward-looking behavior. To that end, researchers can exploit the close connection between plan-ahead behavior and many classical network optimization problems in operations research, as briefly outlined in the following example taken from Hui et al. (2008).

Consider a grocery shopper who carries a shopping list. Assuming that he wants to take the shortest path that allows him to gather his purchases, we can say that the shopper is solving a traveling salesman problem (TSP) (Lawler 1985). Because the TSP is a complex combinatoric problem, this specification clearly assumes too much knowledge and computational ability on the part of the consumer. As a result, a model that assumes consumers follow the path suggested by the optimal solution of the corresponding TSP will be a poor fit to the actual data.

To develop a more realistic descriptive model, researchers may consider two different relaxations that correspond to two different extant research areas in operations research. First, if we assume that a shopper obtains information about the product placements as he moves around, the problem will be isomorphic to a variant of the TSP known as the *online TSP* (e.g., Ausiello et al. 2001), where information about the environment is *released* based on a certain schedule. Second, if we reduce the computational capabilities on the part of the consumer, we may capture planning behavior by a greedy algorithm (i.e., he moves towards the location that carries the item on his shopping list that is not yet purchased and is closest to his current location), or more generally by a greedy algorithm with a certain number of look-ahead steps (Cormen et al. 2001). These extensions allow researchers to borrow methods from the operations research literature to incorporate different degrees/aspects of forward-looking behavior when building models of paths.

3. Literature Review

Based on the framework developed in §2, we provide a literature review for path data that arise in four different areas in marketing: retail/service environments (§3.1), advertising studies (§3.2), e-commerce (§3.3), and experimental research (§3.4). We also briefly outline some path-related areas that are outside of marketing in §3.5.

3.1. Paths in Retail/Service Environments

Marketing researchers have tracked consumers' movements in various retail settings, including grocery stores (Heller 1988), shopping malls (Underhill 2004), and museums (e.g., http://www.crowddynamics.com). The earliest study is Farley and Ring (1966), who physically followed grocery shoppers and documented their movement patterns. Since then, advances in data collection technology have provided researchers with an increasingly sophisticated set of tools to track consumers' movements. Hidden cameras and motion sensors are used to monitor traffic flow (Gogoi 2005), and even capture consumers' eye movements and facial expressions (Pereira 2005). Researchers have also built computer-simulated grocery stores to study people's shopping patterns (Burke 1996, Vrechopoulos et al. 2004). More recently, RFID technology has been used to monitor in-store traffic (Hui et al. 2007, Sorensen 2003). Other promising tracking technologies include the use of global positioning systems and portable shopping devices. Interested readers are encouraged to see Burke (2005) for a comprehensive overview of these new technologies.

To date, researchers have performed various exploratory analyses on shopping paths Heller (1988), measured the flow of shoppers to each department of a grocery store, tabulated the pattern of browsing and purchasing at each product aisle Underhill (2004), qualitatively documented how shoppers interacted with the store environment, and made recommendations to improve shopping convenience. Otnes and McGrath (2001) tracked the shopping paths of male shoppers to test hypotheses about gender-specific shopping behavior. Based on RFID tracking technology, Sorensen (2003) developed a shopper-tracking system known as PathTracker®; his firm has collected and analyzed over 200,000 shopping paths. Larson et al. (2005) later performed a K-medoids clustering algorithm on a sample of these path data and identified a total of 14 different patterns (clusters).

The above research focused on analyzing paths that occur within a certain environment, such as a grocery store or a shopping mall. By contrast, researchers in trip chaining (e.g., Thill and Thomas 1987) studied the paths that people take to perform a set of errands in different geographical locations. Researchers have studied how people select the order of visiting those different destinations, and how they choose the mode of transportation in each trip segment (Adler and Ben-Akiva 1979). For example, Dellaert et al. (1998) investigated how consumers sequence shopping trips that involve three separate destinations (grocery, drugstore, and clothing store). See Thill and Thomas (1987) for an overview of research on trip-chaining behavior.

3.2. Paths in Advertising Studies

In advertising studies, infrared (or near-infrared) corneal reflection technology is often used to trace a subject's eye movements (Duchovski 2002). During the study, an infrared beam, invisible to the eye, is reflected off the cornea, and its angle of reflection is continuously monitored. Because the visual focus of the eye at any moment can then be calculated, the technique is often referred to as eye tracking.

In psychology, eye tracking has been widely used to study reading and information processing patterns (Rayner 1998). Marketing researchers have applied eye-tracking methodologies to capture consumers' eye movements when they view print advertisements (e.g., Janiszewski 1993; Krugman et al. 1994; Pieters et al. 1996, 1999, 2002; Rayner et al. 2001; Rosbergen et al. 1997). For instance, Pieters and Wedel (2004) followed consumers' visual focus and used that as a proxy for the amount of attention paid to the different elements (brand, pictorial, and text) of an advertisement. They then studied how different sizes of each element enhance or reduce overall visual attention. Fox et al. (1998) studied the effectiveness of different alcohol and cigarette warning messages by tracking the amount of time consumers' focus (visually) on the messages and offered practical managerial insights to improve the effectiveness of health warnings.

Marketing researchers have also tracked consumers' visual movements when they are reading yellow page advertising (Lohse 1997), searching the Web for information (Goldberg et al. 2002), and while making brand choice decisions (Chandon et al. 2001, Pieters and Warlop 1999). Other researchers have used eye tracking techniques to study the effect of flash banners (Day et al. 2006), to assess the influence of page layout on visual search patterns (Janiszewski 1998), and even to infer the acceptability of brand extensions (Stewart et al. 2004).

3.3. Paths in E-Commerce

In e-commerce, marketing researchers collect Web browsing (or clickstream) data, which contain detailed click-to-click page-viewing information for each consumer (Bucklin et al. 2002). Montgomery et al. (2004) used Web browsing data, together with a dynamic multinomial probit model that captures page-to-page transitions, to predict purchase conversion. They reported that their model performed better than traditional models that failed to take full advantage of the richness of the path data. Likewise, Moe et al. (2002) and Sismeiro and Bucklin (2004) predicted consumers' online purchasing behavior more accurately by using browsing characteristics as covariates.

The potential applications of Web browsing data analysis extend beyond the prediction of purchase behavior. Researchers also study Web surfing patterns to understand information search behavior and to classify Web browsing strategies. For instance, Chi et al. (2001) developed the idea of an "information scent" to infer a user's information needs given his pattern of surfing. Moe (2003) defined and tabulated 14 different summary statistics of each page-to-page viewing session and used them to classify a visit as buying, browsing, searching or information-building. Similarly, Catledge and Pitkow (1995) recorded user behavior to understand navigational strategies and made suggestions on Web design to enhance usability. Other marketing applications and research opportunities are outlined in Bucklin et al. (2002).

Several advances in Web browsing data analysis have also been made in computer science, and could be applied in marketing in the future. For instance, researchers have developed new data-mining methods to uncover common path traversal patterns to predict a browser's next choice of Web page given his viewing history (Li et al. 2004). Other researchers clustered Web users based on shared interests inferred from their sequence of Web page visits and hyperlink selections (Shahabi et al. 1997), and subsequently profiled consumers based on navigation patterns (e.g., Banerjee and Ghosh 2000, 2001). Computer scientists have patterns from Web browsing data to customize Web page content (Mobasher et al. 2001) and to generate product recommendations using collaborative filtering (e.g., Kim et al. 2004). These and several other applications are discussed in Theusinger and Huber (2000).

3.4. Paths in Experimental Research

Path data arise naturally in experimental settings such as IA (Urban et al. 1997) and Mouselab (Payne et al. 1988). IA belongs to a broad class of multimedia-based testing techniques (Hoffman and Novak 1996) and is often used to predict the success or failure of a new product before it is fully developed. During an IA session, consumers are first accelerated into a future context that conditions them to understand the environment surrounding the new product being tested (e.g., a world with a severe gas shortage). They are then instructed to gather information to decide whether/when to purchase the new product (e.g., an electric automobile). They learn about the new product by viewing, in any order and frequency they choose, a set of simulated product information, such as product attributes, product photos, virtual showroom visits, videos, advertising, newspaper articles, and (simulated) consumer reviews. The path of information viewing and time durations are recorded by the researcher to study consumers' information acquisition and processing patterns (Hauser et al. 1993). Very often an experimental design manipulates the kinds of information seen by different consumers (e.g., some see positive WOM testimonials, while others see negative ones), and researchers can observe the impact on consumers' subsequent information collection (and product purchase) decisions.

Another example that involves path data is Mouselab (e.g., Payne et al. 1988, Sen and Johnson 1997), an early and innovative computer system used by psychologists and marketers to study the process of how a person acquires information and makes decisions. With the Mouselab software, researchers can trace a subject's information acquisition movement by recording the movement of the mouse as a subject answers a series of questions. The resulting data reflect the path of the mouse pointer over time. Other related examples include the use of virtual reality systems that allow users to interactively explore simulated environments, usually to learn about a product (e.g., Volkswagen cars, Sprint telephones) (Lurie and Mason 2007).

Although a system such as Mouselab can generate very interesting path data, we have never seen it used in such a manner. Past researchers have tended to rely on more traditional analyses, e.g., two-sample *t*-tests and aggregate summary statistics. Well-specified models built on the entire Mouselab path data may offer additional insights about observed behavior and the cognitive underlying processes.

3.5. Paths Outside of Marketing

Many researchers outside of marketing have collected and analyzed path data. Researchers have studied the movement of animals (e.g., Polovina et al. 2000, Preisler et al. 2004), insects (e.g., Crist et al. 1992, Jeanson et al. 2003), and pedestrians (e.g., Helbing and Molnar 1995, Teknomo et al. 2000). Reynolds (1987) modeled the motion of a flock of birds using simple behavioral assumptions. He assumed that each bird moves according to three simple rules: (1) it attempts to move closer to the other birds in its neighborhood; (2) it matches the velocity of its flockmates; (3) it avoids collision with other birds. Using computer simulations, Reynolds (1987) replicated realistic flock motions.

Similarly, Helbing and Molnar (1995) proposed a model based on an analogy between people's movement and the principles of Newtonian mechanics. They defined and modeled latent social forces that can be exerted on an individual. In their model, social forces that act on a person may include the attractive force exerted by one's destination, the repellent force from other pedestrians to avoid collision, and the attractive forces caused by the inherent comfort of the ground. The aggregation of these forces in turn governs the direction of a pedestrian's movement.

While some of these concepts may not apply directly to many consumer settings, they provide thought-provoking metaphors that help to derive our framework and offer new ways to think about patterns of shopping, information search, and e-commerce.

4. Issues in Building Models of Path Behavior

In this section, we discuss some modeling issues/choices that researchers will frequently face when developing path models. We outline these issues in parallel with the structure of our framework by first discussing issues for the spatial configuration, the agent, and finally, the paths.

4.1. Modeling the Spatial Configuration

Modeling a spatial configuration as a discretized grid, as compared to a continuous space, can lead to several advantages. First, discretization simplifies the space that may originally contain an infinite number of locations and thus reduces computational complexity (see the two examples in Figure 3). When the sample size is large (e.g., Sorensen 2003 collected 200,000 grocery shopping paths; Larson et al. 2005 analyzed a sample of 27,000 paths), this reduction in complexity may be crucial to enable implementation. Second, because a consumer faces a finite number of alternatives at each step, a discrete spatial configuration ties well with discrete choice models (e.g., Guadagni and Little 1983) used widely in marketing. By contrast, with a continuous space, researchers will need to develop/use a different class of choice models that can handle an infinite number of alternatives. Third, if the discretization is carefully made based on substantive insights, such models will allow for easier interpretation. For instance, if a grocery store is discretized based on the locations of product categories, managers can readily interpret and understand the output by relating them to category-to-category transitions. On the other hand, a continuous model will probably not be as directly linked to substantive insights or managerial actions.

While discretization leads to many advantages in computation, model building, and interpretation, it does have some limitations. First, the process of discretization requires one to recode the original continuous (x, y) data into a discretized form. This data preparation step may be very time consuming if the data set is large or the spatial environment is complex. Second, discretization requires careful construction of a graph, based on substantive insights, to represent the underlying spatial configuration. In problems where little domain knowledge exists, researchers may find it difficult to construct a reasonable representation. A related concern is that the conclusions derived from the path model may vary based on the chosen method of discretization. Thus, we recommend that researchers experiment with different discretization schemes to see how sensitive model conclusions are to different specifications. Finally, the process of discretization reduces the resolution of the data. Because we are approximating continuous data using a discrete grid, variations in the original data at a level finer than the grid will be lost. Thus, researchers may also want to experiment with different levels of granularity in the discretization to determine the optimal resolution.

4.2. Modeling the Agent

To model the characteristics of the agent, two particularly important research issues need to be considered. First, researchers need to specify the role and the nature of the heterogeneity among agents. Second, because an agent's behavior may change over the course of his path, adaptive/nonstationary behaviors should also be taken into account.

4.2.1. Role(s) of Heterogeneity. Researchers need to specify the dimension(s) on which consumers are considered to be heterogeneous to calibrate their path models. This requires careful thought because consumers can be heterogeneous in many different aspects, including all of the dimensions we have discussed, i.e., social interactions, goals (if any), and forward-looking tendencies.

For instance, consumers may exhibit different responses to social interaction; researchers have documented that even given the same in-store shopper density, people may have different perceptions of how crowded the area is. Perceived crowding is influenced by individual characteristics and situational factors (Harrell and Hutt 1976). In addition, consumers may respond differently to crowding situations: while some may shorten the time they spend choosing between products, others may become more goal-directed and spend less effort in exploration.

Likewise, consumers may enter the spatial configuration with different goals. In a grocery store, for example, each consumer may come in with a different set of product categories that he wants to purchase; some shoppers, who may be hedonic browsers or window-shoppers, may not have a clear set of purchase goals. Researchers may consider capturing goal and social-interaction heterogeneity by specifying a set of individual-level parameters that describe such behavior and linking them via a Hierarchical Bayes framework (Rossi et al. 2006), or by segmenting consumers into different classes, based on their social behavior and goal-directedness, using a latent class model (Kamakura and Russell 1989).

Accounting for heterogeneity in consumers' forward-looking behavior requires even greater thought because of the difficulty in representing forward-looking behavior as a parameter. As discussed in §2.2.3, researchers may exploit the close connection between planning behavior and many classical network optimization problems (Cormen et al. 2001). This connection may allow researchers to experiment with a wide array of model specifications based on different assumptions about consumers' knowledge of the environment (e.g., Moorman et al. 2004), their degree of rationality, and the optimization algorithm they are presumed to follow. Researchers can then test these model specifications with field data from a panel of consumers with different levels of knowledge about the spatial environment. For instance, based on an analogy with the TSP, Hui et al. (2008) analyzed the efficiencies of grocery shoppers to infer the number of look-ahead steps, similar to Camerer and Ho (1999) in their study of behavior in experimental economics games.

4.2.2. Adaptive/Nonstationarity Behavior. Researchers also need to take into account the fact that a path process may not necessarily be stationary. That is, as consumers progress in a trip, they may exhibit some degree of adaptive/nonstationary behavior. First, some changes may occur naturally as part of the path process. For example, a Web user may naturally switch from an initial browsing state, when he is casually surfing the Web for general information, to a deliberation state, when he decides that he wants to purchase certain products (Montgomery et al. 2004). Similarly, Bates (1989) hypothesized a new model of information search called "berrypicking," which assumed that users obtain information "a bit at a time," and that their information needs evolve over time based on previous search results. In both cases, the consumers' goals have changed over time and must be carefully modeled.

Second, some changes may accumulate gradually over the path. For instance, a grocery shopper may become more fatigued the longer she spends in the store. Thus, we may observe that as time progresses, the shopper may want to spend less time exploring the store and instead use a more directed approach to pick up only those products that she needs. Indeed, Sorensen (2003) documented that shoppers tend to speed up the more time they spent in the store.

Finally, some changes may also be driven by social influence, as consumers adapt their behavior based on what other consumers do, e.g., to avoid crowding or to preserve interpersonal space (Underhill 2000).

These different types and sources of nonstationarity are important because they may necessitate different modeling methodologies. For sudden changes that involve a discontinuous change of state, approaches such as hidden Markov models (Rabiner 1989), or more generally state-space models (Kim and Nelson 1999), where consumer behavior is modeled conditional on their hidden states (which may evolve over time through some stochastic processes), may be more appropriate. If the changes are gradual, however, it may be more appropriate to incorporate the continuous change as a parameter in the model and allow for a trend component on that parameter to capture its smooth time-varying nature.

4.3. Analysis and Modeling of Paths

We now discuss some issues that arise in the analysis and modeling of paths. First, we describe some statistical issues involved in exploratory analyses of paths. Second, we discuss the deterministic/stochastic nature of movements. Finally, we explore the challenges in building models that integrate data collected from paths and other behaviors (e.g., purchases).

4.3.1. Statistical Issues for Exploratory Analyses. Before developing a formal model of paths, it is helpful for a researcher to graphically display the data and perform exploratory analyses (Wainer 2004). Performing exploratory analyses on paths, however, is nontrivial and can pose major challenges for marketing modelers and statisticians. By nature, path data are multivariate and therefore relatively complicated: each record in a path data set is a multivariate sequence that represents the consumer's position over time. In addition, path data are fundamentally different from traditional multivariate data because, as discussed in our definition of paths, there is a continuous relationship between each position over time. Commonly used multivariate techniques (e.g., cluster analysis, factor analysis) that ignore this temporal relationship may not be suitable. Instead, statisticians and modelers need to consider a number of key statistical issues, such as the following:

(1) How should we display and compare paths that are different in sampling intervals? For example, some paths may contain record movements in 5-second intervals, while others in 15-second intervals. How should we store and format these kinds of data?

(2) How should we compare paths of different durations? For instance, a shopper who spends two hours at a grocery store will have a much longer path than a shopper who spends five minutes. How do we create a metric/topology of paths to calculate the distance between these paths?

(3) How can we summarize the sources of variation among paths, taking into account constraints within the spatial configuration (e.g., walls, aisles)?

Larson et al. (2005) approached this problem using a *K*-medoid clustering technique applied to grocery cart movement data. It would be interesting if, in addition to obtaining a clustering solution, researchers could describe path data using ideas from principal components analysis (PCA) to break each path into some combination of lower-level components. For instance, Bradlow (2002) described an approach to apply PCA to repeated measures data sets to explore their key features. With some modifications, researchers might similarly extend PCA methods to extract key information from path data.

More generally, functional data analysis (Ramsay and Silverman 2005) seems to provide another appropriate set of tools to address these issues. In functional data analysis, the underlying data are functions themselves; path data are functions of time, with the function value at each time point being the agent's location. Although many sophisticated exploratory techniques have been developed in functional data analysis, they cannot be directly applied to path data. This is mainly because existing methodology typically assumes that the underlying space is unrestricted, while path data often contain spatial constraints. Identifying a proper method to handle these constraints is a crucial issue for those who wish to apply functional data analysis techniques to path data.

Another major challenge in exploring path data is how one should take into account other ancillary information available for each path. For example, a data set of grocery shopping paths not only includes the shopping path information but also typically will be linked to purchases. Because the data come in two different forms, a path (functional data) and a record of purchases (multivariate data), existing exploratory techniques (e.g., clustering) would be ill-suited to describe the relationship. One promising direction is the use of a mixed-PCA algorithm developed in functional data analysis. This may allow researchers to analyze data sets containing both functional and multivariate data (Ramsay and Silverman 2005).

4.3.2. Deterministic vs. Stochastic. The law of motion for a given model, i.e., the exact mechanism with which movements are determined given all the characteristics of the agents and the spatial configuration, can be deterministic or stochastic. In areas outside of marketing, researchers commonly use a deterministic specification to model the movement of birds and pedestrians. Given an initial configuration of birds and their migration target, the path of the flock is generally assumed to be completely determined (Reynolds 1987). In the social force model for pedestrians, a potential surface is calculated from the vector summation of all social forces acting on a pedestrian, which reflects the attractiveness of each location (Helbing et al. 1997). Given this potential surface, the pedestrian is assumed to move in the direction with the largest increase in ground attraction, much like how particles move in a force field.

In marketing, by contrast, a stochastic framework is often used to model movements. For instance, Montgomery et al. (2004)'s Web browsing model specifies that the utility of a category is a linear function of covariates and a vector-autoregressive (VAR) structure, plus a (multivariate) normally distributed error term. These types of components are commonly used in random utility models used in econometrics; it allows for the nondeterministic nature of click-to-click level Web browsing patterns, even after accounting for all observed and unobserved covariates. Likewise, models of eye tracking scanpaths often assume that the path of visual movement is stochastic, and is usually modeled (if the space is discretized) using multivariate logit-type choice models (Pieters and Wedel 2004).

The fusion between deterministic and stochastic formulations leads to many intriguing opportunities for modeling. While a stochastic formulation is more flexible and consistent with random utility models in marketing, we believe that the introduction of some deterministic components within a general stochastic formulation may allow for a richer and more structural model of movement. For instance, researchers have documented that pedestrians tend to make right turns (Bitgood and Dukes 2006) and tend to speed up as they progress in a shopping trip (Sorensen 2003). To incorporate these and other general behavioral tendencies, we suggest that they can be modeled as deterministic components in a path model, while finer-level decisions can be modeled using stochastic terms.

4.3.3. Degree of Integration for Multivariate Observed Behaviors. A random-utility specification for a path model may facilitate the fusion between path models and other discrete choice models (e.g., purchase/brand choice). Such fusion has not been a hallmark of modeling work in this area and would be a significant contribution. In other domains, however, marketers have built integrated models of purchase incidence, brand choice, and purchase quantity (Chintagunta 1993), as well as the "who, when, where, and how much" characteristics of auctions (Park and Bradlow 2005). The integration of movement and choice data would be in the same spirit and offers many intriguing research opportunities.

One key issue to consider in this promising direction is the degree of integration for the observed behaviors. An integration of path and other behaviors only makes sense if they are strongly connected. This may hold in some settings. For example, one would assume that in a grocery store, a person will be more likely to visit the product aisle that he intends to buy from. Thus, the final market basket probably has a strong relationship with the path that the shopper takes. In other settings, however, the connection between one's path and other behaviors may not be evident. For instance, a person who spends a long time looking for information about several books on Amazon.com may not actually buy from the website, but may be using it instead to gather information about a particular author. In a model of online conversion behavior, Moe and Fader (2004) found a considerable number of these "hard-core non-buyers" in the Web browsing data from this website. In such a case, the connection between the path taken and purchases can be weak.

To study the degree of integration between paths and other activities, researchers may consider fitting two (nested) models: one that allows for the integration between the path and other activities (e.g., purchase), and the other that explicitly turns off this linkage by restricting some of the parameter values. The relative fit of the nested models allows researchers to understand whether a model that integrates different behaviors is superior. With more such exercises in the future, researchers will be able to make some empirical generalizations about the strength of the connection between different kinds of activities and the consumer's path to achieve them.

5. Conclusion and Future Research Directions

The main objective of this paper is to motivate and clarify research issues related to path data in marketing. Understanding the interaction between a consumer's behavior and spatial environment allows us to extract vital insights about how and why consumers engage in certain behaviors. Using grocery shopping, Web browsing, eye tracking, and IA as primary examples, we find that marketing researchers have long been collecting and analyzing path data, but that there has never been a general framework linking them. We outline a number of different areas where experimental and empirical researchers may collect and analyze path data. We then synthesize these different areas by highlighting the role of paths in a unifying framework to guide future research in this broad area.

We offer an integrative framework of path models that pinpoints their various components and dimensions. We identified the two key components of a path model: the spatial configuration and the agent. The spatial configuration can be classified along three subdimensions: (1) physical/nonphysical, (2) continuous/discrete, and (3) the degree of constraints. Agents can be characterized based on (1) their degree of social interaction, (2) their degree of goal-directedness, and (3) their forward-looking tendencies. We then discuss an array of statistical and modeling issues that researchers might face in constructing a path model. While we do not claim to provide a complete list of challenges and issues, we hope to offer some useful guidance to future researchers faced with path-related data sets.

As path data sets become more commonly available, we expect to see more research activities in this new and promising area. These activities will not only include descriptive models of movements and choices (as highlighted in this paper), but also a richer variety of research issues. Below, we briefly describe three possible areas that reflect the kinds of investigations we hope to see in the near future.

(1) *Hierarchies of paths*: In this paper, we focused on agents operating in a single spatial environment. More generally, one may consider a hierarchy of paths. At a high level, a consumer decides the sequence of websites to visit (e.g., Park and Fader 2004). Within each website, he may then navigate among the Web pages with a specific click-to-click pattern (e.g., Montgomery et al. 2004). Similar issues arise in a shopping mall or even a large store (Arentze et al. 2005, Brooks et al. 2004), where there may be a hierarchy of paths in effect (i.e., deciding which stores/departments to visit, then beginning a new path within each one). Observations about the paths within limited parts of the hierarchy would be highly informative about the shopper's behavior in other parts of it.

(2) *Inferences about partially-observed paths*: As path research matures, we may find that we do not need all of the movement data to make accurate statements about path-related behaviors. For instance, we may need only a series of still images from key locations in the environment to capture the relevant aspects of the path, which could enable significant cost savings. To assess this possibility, one can fit two models to the data—one with only location data (e.g., which aisles a shopper visits), and the other with the full path data (i.e., the sequence of aisle visits). This would allow researchers to assess the loss of information in throwing away much of the movement data. Research firms such as Sorensen Associates are actively investigating such methods, and modeling techniques such as those developed by Musalem et al. (2008, 2009) make this a promising area of academic research.

(3) *Convergence of different types of paths*: Although we have carefully compared and contrasted various types of marketing paths throughout the paper, we have treated each one as a separate entity, without considering possible links among them. As data collection technologies become more sophisticated and economical, however, we envision a possible convergence of different path domains; i.e., researchers may simultaneously collect and combine different data sources. Several current sources of convergence include: (1) companies that not only track typical page-to-page Web browsing data, but also the movie of mouse movements within each page to see how users interact with a website (http://www.clicktale.com); (2) touch-screen kiosks in retail stores (or even on the store window—see http://wcbstv.com/technology/local_story_219114456.html), which combine elements of Web browsing and in-store path data; and (3) collecting eye tracking data at the store shelf, which can be combined with customer movement data. All of these data sources (and countless other areas where convergence of path data might occur) will create

more exciting challenges and opportunities for marketing researchers who wish to learn even more about different aspects of consumer behavior.

We encourage other researchers to join us in the study of paths to investigate these and other interesting issues.

Acknowledgments

The authors thank Herb Sorensen of TNS-Sorensen for his feedback and encouragement.

References

Adler, T., M. Ben-Akiva. 1979. A theoretical and empirical model of trip chaining behavior. *Transportation Res. B* **13** 243–257.

Arentze, T. A., H. Oppewal, H. Timmermans. 2005. A multipurpose shopping trip model to assess retail agglomeration effects. *J. Marketing Res.* **42**(1) 109–115.

Argo, J. J., D. W. Dahl, R. V. Manchanda. 2005. The influence of a mere social presence in a retail context. *J. Consumer Res.* **32**(2) 207–212.

Ausiello, G., E. Feuerstein, S. Leonardi, L. Stougie, M. Talamo. 2001. Algorithms for the on-line traveling salesman. *Algorithmica* **29**(4) 560–581.

Banerjee, A. V. 1992. A simple model of herd behavior. *Quart. J. Econom.* **107**(3) 797–817.

Banerjee, A., J. Ghosh. 2000. Concept-based clustering of clickstream data. *Proc. 3rd Internat. Conf. Inform. Tech.*, Bhubaneswar, India, 145–150.

Banerjee, A., J. Ghosh. 2001. Clickstream clustering using weighted longest common subsequences. *Proc. SIAM Conf. Data Mining: Workshop Web Mining, Chicago*, ACM, New York, 33–40.

Banerjee, S., B. P. Carlin, A. E. Gelfand. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall, Boca Raton, FL.

Bates, M. J. 1989. The design of browsing and berrypicking techniques for the online search interface. *Online Rev.* **13**(5) 407–424.

Becker, G. S. 1991. A note on restaurant pricing and other examples of social influence on price. *J. Political Econom.* **99**(5) 1109–1116.

Bell, D., S. Song. 2007. Neighborhood effects and trial on the Internet: Evidence from online grocery retailing. *Quant. Marketing Econom.* **5**(4) 361–400.

Bitgood, S., S. Dukes. 2006. Not another step! Economy of movement and pedestrian choice point behavior in shopping malls. *Environ. Behav.* **38**(3) 394–405.

Blackmore, S. 2003. *Consciousness*: *An Introduction*. Oxford University Press, New York.

Bollobas, B. 1979. *Graph Theory*: *An Introductory Course*. Springer-Verlag, New York.

Bradlow, E. T. 2002. Exploring repeated measure data sets for key features using principal components analysis. *Internat. J. Res. Marketing* **19**(2) 167–179.

Bradlow, E. T., B. Bronnenberg, G. J. Russell, N. Arora, D. R. Bell, S. D. Duvvuri, F. T. Hofstede, C. Sismeiro, R. Thomadsen, S. Yang. 2005. Spatial models in marketing. *Marketing Lett.* **16**(3–4) 267–678.

Bril, G. 1995. Forecasting hurricane tracks using the Kalman filter. *Environmetrics* **6**(1) 7–16.

Broder, A., R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener. 2000. Graph structure of the Web. *Comput. Networks* **33**(1–6) 309–320.

Brooks, C. M., P. J. Kaufmann, D. R. Lichtenstein. 2004. Travel configuration on consumer trip-chained store choice. *J. Consumer Res.* **31**(2) 241–248.

Bucklin, R. E., J. M. Lattin, A. Ansari, S. Gupta, D. Bell, E. Coupey, J. D. C. Little, C. Mela, A. Montgomery, J. Steckel. 2002. Choice and the Internet: From Clickstream to research stream. *Marketing Lett.* **13**(3) 245–258.

Burke, R. R. 1996. Virtual shopping: Breakthrough in marketing research. *Harvard Bus. Rev.* **74**(2) 120–131.

Burke, R. R. 2005. The third wave of marketing intelligence. M. Drafft, M. Mantrala, eds. *Retailing in the 21st Century*: *Current and Future Trends*. Springer, New York, 113–125.

Camerer, C., T.-H. Ho. 1999. Experience-weighted attraction learning in normal form games. *Econometrica* **67**(4) 827–874.

Catledge, L. D., J. E. Pitkow. 1995. Characterizing browsing strategies in the World-Wide Web. *Comput. Networks ISDN Systems* **27**(6) 1065–1073.

Chandon, P., J. W. Hutchison, E. Bradlow, S. H. Young. 2006. Measuring the value of point-of-purchase marketing with commercial eye-tracking data. INSEAD Business School Reasearch Paper 2007/22/MKT/ACGRD, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1032162.

Chi, E. H., P. Pirolli, K. Chen, J. Pitkow. 2001. Using information scent to model user information needs and actions on the Web. *Proc. Human Factors Comput. Systems (CHI 2001)*, ACM Special Interest Group on Computer-Human Interactions, Seattle, 490–497.

Chintagunta, P. K. 1993. Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Sci.* **12**(2) 184–208.

Chopard, B., P. Luthi, P.-A. Queloz. 1996. Cellular automata model of car traffic in two-dimensional street networks. *J. Phys. A* **29** 2325–2336.

Cormen, T. H., C. E. Leiserson, R. L. Rivest, C. Stein. 2001. *Introduction to Algorithms*, 2nd ed. MIT Press, Cambridge, MA.

Crist, T. O., D. S. Guertin, J. A. Wiens, B. T. Milne. 1992. Animal movement in heterogeneous landscapes: An experiment with Eleodes beetles in shortgrass prairie. *Funct. Ecology* **6**(5) 536–544.

Day, R.-F., G. C. W. Shyi, J.-C. Wang. 2006. The effect of flash banners on multiattribute decision making: Distractor or source of arousal? *Psych. Marketing* **23**(5) 369–382.

Dellaert, B. G., T. A. Arentze, M. Bierlaire, A. W. J. Borgers, H. J. P. Timmermans. 1998. Investigating consumers' tendency to combine multiple shopping purposes and destinations. *J. Marketing Res.* **35**(2) 177–188.

Du, R. Y., W. A. Kamakura. 2006. Household life cycles and lifestyles in the United States. *J. Marketing Res.* **43**(1) 121–132.

Duchovski, A. T. 2002. A breadth-first survey of eye tracking applications. *Behav. Res. Methods, Instruments, Comput.* **34**(4) 455–470.

Eirinaki, M., M. Vazirgiannis, D. Kapogiannis. 2005. Web path recommendations based on page ranking and Markov models. *Proc. 7th Annual ACM Internat. Workshop*, ACM, New York, 2–9.

Farley, J. U., L. W. Ring. 1966. A stochastic model of supermarket traffic flow. *Oper. Res.* **14**(4) 555–567.

Fox, R. J., D. M. Krugman, J. E. Fletcher, P. M. Fischer. 1998. Adolescents' attention to beer and cigarette print ads and associated product warnings. *J. Advertising Res.* **27**(3) 57–68.

Gogoi, P. 2005. Retailing, the high-tech way. *BusinessWeek Online* (July 6) Special Report, Retailing's New Tech.

Goldberg, J. H., M. J. Stimson, M. Lewenstein, N. Scott, A. M. Wichansky. 2002. Eye tracking in Web search tasks: Design implications. *Proc. Sympos. Eye Tracking Res. Appl., New Orleans*, ACM Press, New York 51–58.

Guadagni, P. M., J. D. C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Sci.* **2**(3) 203–238.

Harrell, G. D., M. D. Hutt. 1976. Buyer behavior under conditions of crowding: An initial framework. *Adv. Consumer Res.* **3** 36–39.

Harrell, G. D., M. D. Hutt, J. C. Anderson. 1980. Path analysis of buyer behavior under conditions of crowding. *J. Marketing Res.* **17**(1) 45–51.

Hauser, J. R., G. L. Urban, B. D. Weinberg. 1993. How consumers allocate their time when searching for information. *J. Marketing Res.* **15**(4) 452–466.

Helbing, D., P. Molnar. 1995. Social force model for pedestrian dynamics. *Phys. Rev. E* **51**(5) 4282–4286.

Helbing, D., J. Keltsch, P. Molnar. 1997. Modelling the evolution of human trail systems. *Nature* **388**(6637) 47–49.

Heller, W. 1988. Tracking shoppers through the combination store. *Progressive Grocer* (November) 47–64.

Henderson, L. F. 1974. On the fluid mechanics of human crowd motion. *Transportation Res.* **8** 509–515.

Hoffman, D. L., T. P. Novak. 1996. Marketing in hypermedia computer-mediated environments: Conceptual foundations. *J. Marketing* **60** 50–68.

Hui, S. K., E. T. Bradlow, P. S. Fader. 2007. An integrated model of grocery store shopping path and purchase behavior. Working paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=960960.

Hui, S. K., P. S. Fader, E. T. Bradlow. 2008. The traveling salesman goes shopping: The systematic deviations of grocery paths from TSP optimality. *Marketing Sci.* ePub ahead of print October 9, http://mktsci.journal.informs.org/cgi/content/abstract/mksc.1080.0402v1.

Janiszewski, C. 1993. Preattentive mere exposure effects. *J. Consumer Res.* **20**(3) 376–392.

Janiszewski, C. 1998. The influence of display characteristics on visual exploratory search behavior. *J. Consumer Res.* **25**(3) 290–301.

Jeanson, R., S. Blanco, R. Fournier, J.-L. Deneubourg, V. Fourcassie, G. Theraulaz. 2003. A model of animal movements in a bounded space. *J. Theoret. Biol.* **225** 443–451.

Kalawsky, R. S. 1993. *The Science of Virtual Reality and Virtual Environments*: *A Technical, Scientific, and Engineering Reference on Virtual Environments*. Addison-Wesley Wokingham, Wokingham, Berks, UK.

Kamakura, W. A., G. J. Russell. 1989. A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* **26**(4) 379–390.

Kim, C.-J., C. R. Nelson. 1999. *State-Space Models with Regime Switching*: *Classical and Gibbs-Sampling Approaches with Applications*. MIT Press, Cambridge, MA.

Kim, D.-H., V. Atluri, M. Bieber, N. Adam, Y. Yesha. 2004. A clickstream-based collaborative filtering personalization model: Towards a better performance. *Proc. 6th Annual ACM Internat. Workshop Web Inform. Data Management*, ACM, New York, 88–95.

Krugman, D. M., R. J. Fox, J. E. Fletcher, P. M. Fischer, T. H. Rojas. 1994. Do adolescents attend to warnings in cigarette advertising? An eye-tracking approach. *J. Advertising Res.* 39–52.

Larson, J. S., E. T. Bradlow, P. S. Fader. 2005. An exploratory look at supermarket shopping paths. *Internat. J. Res. Marketing* **22**(4) 395–414.

Lawler, E. L. 1985. *The Traveling Salesman Problem*: *A Guided Tour of Combinatorial Optimization*. John Wiley & Sons, New York.

Lee, L., D. Ariely. 2006. Shopping goals, goal concerteness, and conditional promotions. *J. Consumer Res.* **33** 60–70.

Li, H.-F., S.-Y. Lee, M.-K. Shan. 2004. On mining webclick streams for path traversal patterns. *2004 Internat. World Wide Web Conf., Proc. 13th Internat. World Wide Web Conf.*, ACM, New York, 404–405.

Lohse, G. L. 1997. Consumer eye movement patterns on yellow pages advertising. *J. Advertising* **16**(1) 61–73.

Lurie, N., C. H. Mason. 2007. Visual representation: Implications for decision making. *J. Marketing* **71** 160–177.

Mobasher, B., H. Dai, T. Lou, M. Nakagawa. 2001. Effective personalization based on association rule discovery from Web usage data. *Proc. 3rd ACM Workshop Web Inform. Data Management, Atlanta*, ACM, New York, 9–15.

Moe, W. W. 2003. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *J. Consumer Psych.* **13** 29–39.

Moe, W. W., P. S. Fader. 2004. Dynamic conversion behavior at e-commerce sites. *Management Sci.* **50**(3) 326–335.

Moe, W. W., H. Chipman, E. George, R. McCulloch. 2002. A Bayesian treed model of online purchasing behavior using in-store navigational clickstream. Working paper, University of Texas at Austin, Austin.

Montgomery, A. L., S. Li, K. Srinivasan, J. C. Liechty. 2004. Modeling online browsing and path analysis using clickstream data. *Marketing Sci.* **23**(4) 579–595.

Moorman, C., K. Diehl, D. Brinberg, B. Kidwell. 2004. Subjective knowledge, search locations, and consumer choice. *J. Consumer Res.* **31** 673–680.

Musalem, A., E. T. Bradlow, J. S. Raju. 2008. Who's got the coupon? Estimating consumer preferences and coupon usage from aggregate information. *J. Marketing Res.* **45**(6) 715–730.

Musalem, A., E. T. Bradlow, J. S. Raju. 2009. Bayesian estimation of random-coefficients choice models using aggregate data. *J. Appl. Econometrics*. Forthcoming.

Otnes, C., M. A. McGrath. 2001. Perceptions and realities of male shopping behavior. *J. Retailing* **77** 111–137.

Park, Y.-H., E. T. Bradlow. 2005. An integrated model for bidding behavior in Internet auctions: Whether, who, when, and how much. *J. Marketing Res.* **42**(4) 470–482.

Park, Y.-H., P. S. Fader. 2004. Modeling browsing behavior at multiple websites. *Marketing Sci.* **23** 280–303.

Payne, J. W., J. R. Bettman, E. J. Johnson. 1988. Adaptive strategy selection in decision making. *J. Experiment. Psych.* **14**(3) 534–552.

Pereira, J. 2005. Spying on the sales floor. *Wall Street Journal* (December 21) B1.

Pieters, R. G. M., L. Warlop. 1999. Visual attention during brand choice: The impact of time pressure and task motivation. *Internat. J. Res. Marketing* **16**(1) 1–16.

Pieters, R. G. M., M. Wedel. 2004. Attention capture and transfer in advertisting: Brand, pictorial, and text-size effects. *J. Marketing* **68**(April) 36–50.

Pieters, R. G. M., E. Rosbergen, M. Hartog. 1996. Visual attention to advertising: The impact of motivation and repetition. *Adv. Consumer Res.* **23** 242–248.

Pieters, R. G. M., E. Rosbergen, M. Wedel. 1999. Visual attention to repeated print advertising: A test of scanpath theory. *J. Marketing Res.* **18**(November) 424–438.

Pieters, R. G. M., L. Warlop, M. Wedel. 2002. Breaking through the clutter: Benefits of advertising originality and familiarity for brand attention and memory. *Management Sci.* **48**(6) 765–781.

Polovina, J. J., D. R. Kobayashi, D. M. Parker, M. P. Seki, G. H. Balazs. 2000. Turtles on the edge: Movement of loggerhead turtles (*Caretta caretta*) along oceanic fronts, spanning longline fishing grounds in the Central North Pacific, 1997–1998. *Fisheries Oceanography* **9** 71–82.

Preisler, H. K., A. A. Ager, B. K. Johnson, J. G. Kie. 2004. Modeling animal movements using stochastic differential equations. *Environmetrics* **15** 643–657.

Rabiner, L. R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **777**(2) 257–286.

Ramsay, J., B. W. Silverman. 2005. *Functional Data Analysis*. Springer, New York.

Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psych. Bull.* **124**(3) 372–422.

Rayner, K., C. M. Rotello, A. J. Stewart, J. Keir, S. A. Duffy. 2001. Integrating text and pictorial information: Eye movements when looking at print advertisement. *J. Experiment. Psych. Appl.* **7**(3) 219–226.

Reynolds, C. W. 1987. Flocks, herds and schools: A distributed behavioral model. *Comput. Graphics* **21**(4) 25–34.

Robertson, S. E. 1977. Theories and models in information retrieval. *J. Documentation* **33**(2) 126–148.

Rosbergen, E., R. Pieters, M. Wedel. 1997. Visual attention to advertising: A segment-level analysis. *J. Consumer Res.* **24**(3) 305–314.

Rossi, P. E., G. Allenby, R. McCulloch. 2006. *Bayesian Statistics and Marketing*. John Wiley & Sons, New York.

Rudin, W. 1976. *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, New York.

Sen, S., E. J. Johnson. 1997. Mere-possession effects without possession in consumer choice. *J. Consumer Res.* **24**(1) 105–117.

Shahabi, C., A. M. Zarkesh, J. Adibi, V. Shah. 1997. Knowledge discovery from users Web-page navigation. *Proc. IEEE Researcher Issues in Data Engrg.*, Birmingham, UK, 20.

Sismeiro, C., R. E. Bucklin. 2004. Modeling purchase behavior at an e-commerce website: A task completion approach. *J. Marketing Res.* **16** 306–323.

Song, I., P. K. Chintagunta. 2003. A micromodel of new product adoption with heterogeneous and forward-looking consumers: Application to the digital camera category. *Quant. Marketing Econom.* **1**(4) 371–407.

Sorensen, H. 2003. The science of shopping. *Marketing Res.* **15**(3) 30–35.

Stewart, A. J., M. J. Pickering, P. Sturt. 2004. Using eye movements during reading as an implicit measure of the acceptability of brand extensions. *Appl. Cognitive Psych.* **18** 697–709.

Sun, B., S. A. Neslin, K. Srinivasan. 2003. Measuring the impact of promotions on brand switching when consumers are forward-looking. *J. Marketing Res.* **40**(November) 389–405.

Teknomo, K., Y. Takeyama, H. Inamura. 2000. Determination of pedestrian flow performance based on video tracking and microscopic simulations. *Proc. Infrastructure Planning* **23**(1) 639–642.

Theusinger, C., K.-P. Huber. 2000. Analyzing the footsteps of your customers. *Web Knowledge Discovery and Data Mining* (*WEBKDD*), *Boston*. http://robotics.stanford.edu/~ronnyk/WEBKDD2000/papers/theusinger.pdf.

Thill, J.-C., I. Thomas. 1987. Toward conceptualizing trip-chaining behavior: A review. *Geographical Anal.* **19**(1) 1–17.

Thrun, S., W. Burgard, D. Fox. 2005. *Probabilistic Robotics*. MIT Press, Cambridge, MA.

Underhill, P. 2000. *How We Buy*: *The Science of Shopping*. Simon & Schuster, New York.

Underhill, P. 2004. *Call of the Mall*: *The Geography of Shopping*. Simon & Schuster, New York.

Urban, G. L., J. R. Hauser, W. J. Qualls, B. D. Weinberg, J. D. Bohlmann, R. A. Chicos. 1997. Information acceleration: Validation and lessons from the field. *J. Marketing Res.* **34**(1) 143–153.

Vrechopoulos, A. P., R. M. O'Keefe, G. I. Doukidis, G. J. Siomkos. 2004. Virtual store layout: An experimental comparison in the context of grocery retail. *J. Retailing* **80** 13–22.

Wainer, H. 2004. *Graphic Discovery*: *A Trout in the Milk and Other Visual Adventures*. Princeton University Press, Princeton, NJ.

Zienkiewicz, O. C., R. L. Taylor, J. Z. Zhu. 2005. *The Finite Element Method*: *Its Basis and Fundamentals*. Elsevier Butterworth-Heinemann, Oxford, UK.