Emerald Books

Review of Marketing Research
Volume 4

Review of Marketing Research

Naresh K. Malhotra
Editor

Emerald

REVIEW OF MARKETING RESEARCH

REVIEW OF MARKETING RESEARCH

VOLUME 4

NARESH K. MALHOTRA
EDITOR

United Kingdom – North America – Japan
India – Malaysia – China

Emerald Group Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

Reprints and permission service
Contact: books@emeraldinsight.com

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-7656-2092-7

Awarded in recognition of Emerald's production department's adherence to quality systems and processes when preparing scholarly journals for print

INVESTOR IN PEOPLE

REVIEW OF MARKETING RESEARCH

AD HOC REVIEWERS

Anocha Aribarg, University of Michigan
Ulf Bockenholt, McGill University
Eric Bradlow, University of Pennsylvania
Michael Kamins, University of Southern California
Deborah MacInnis, University of Southern California
Joseph Pancras, University of Connecticut
C. W. Park, University of Southern California
S. Sriram, University of Connecticut
Rajkumar Venkatesan, University of Virginia
Gal Zauberman, University of North Carolina, Chapel Hill

CONTENTS

REVIEW OF MARKETING RESEARCH

Taking Stock

NARESH K. MALHOTRA

Overview

Review of Marketing Research, now in its fourth volume, is a recent publication covering the important areas of marketing research with a more comprehensive state-of-the-art orientation. The chapters in this publication will review the literature in a particular area, offer a critical commentary, develop an innovative framework, and discuss future developments, as well as present specific empirical studies. The first four volumes feature some of the top researchers and scholars in our discipline who have reviewed an array of important topics. The response to the first three volumes has been truly gratifying and we look forward to the impact of the fourth volume with great anticipation.

Publication Mission

The purpose of this series is to provide current, comprehensive, state-of-the-art articles in review of marketing research. Wide-ranging paradigmatic or theoretical, or substantive agendas are appropriate for this publication. This includes a wide range of theoretical perspectives, paradigms, data (qualitative, survey, experimental, ethnographic, secondary, etc.), and topics related to the study and explanation of marketing-related phenomenon. We hope to reflect an eclectic mixture of theory, data, and research methods that is indicative of a publication driven by important theoretical and substantive problems. We seek studies that make important theoretical, substantive, empirical, methodological, measurement, and modeling contributions. Any topic that fits under the broad area of "marketing research" is relevant. In short, our mission is to publish the best reviews in the discipline.

Thus, this publication will bridge the gap left by current marketing research publications. Current marketing research publications such as the *Journal of Marketing Research* (USA), *International Journal of Marketing Research* (UK), and *International Journal of Research in Marketing* (Europe) publish academic articles with a major constraint on the length. In contrast, *Review of Marketing Research* will publish much longer articles that are not only theoretically rigorous but also more expository, with a focus on implementing new marketing research concepts and procedures. This will also serve to distinguish this publication from *Marketing Research* magazine published by the American Marketing Association (AMA).

Articles in *Review of Marketing Research* should address the following issues:

- Critically review the existing literature
- Summarize what we know about the subject—key findings
- Present the main theories and frameworks
- Review and give an exposition of key methodologies
- Identify the gaps in literature
- Present empirical studies (for empirical papers only)
- Discuss emerging trends and issues
- Focus on international developments
- Suggest directions for future theory development and testing
- Recommend guidelines for implementing new procedures and concepts

Articles in the First Volume

The inaugural volume exemplified the broad scope of the *Review of Marketing Research.* It contained a diverse set of review articles covering areas such as emotions, beauty, information search, business and marketing strategy, organizational performance, reference scales, and correspondence analysis. These articles were contributed by some of the leading scholars in the field, five of them being former editors of major journals (*Journal of Marketing* and *Journal of Consumer Research*).

Johnson and Stewart provided a review of traditional approaches to the analysis of emotion in the context of consumer behavior. They reviewed appraisal theory and discussed examples of its application in the contexts of advertising, customer satisfaction, product design, and retail shopping. Holbrook explored and reviewed the concept of beauty as experienced by ordinary consumers in their everyday lives. His typology conceptualized everyday usage of the term "beauty" as falling into eight categories distinguished on the basis of three dichotomies: (i) extrinsically/intrinsically motivated; (ii) thing(s)-/person(s)-based; and (iii) concrete/abstract. Xia and Monroe first reviewed the literature on consumer information search, and then the literature on browsing. They proposed an extended consumer information acquisition framework and outlined relevant substantive and methodological issues for future research. Hunt and Morgan reviewed the progress and prospects of the "resource-advantage" (R-A) theory. They examined in detail the theory's foundational premises, showed how R-A theory provides a theoretical foundation for business and marketing strategy, and discussed the theory's future prospects. Bharadwaj and Varadarajan provided an interdisciplinary review and perspective on the determinants of organizational performance. They examined the classical industrial organization school, the efficiency/revisionist school, the strategic groups school, the business policy school, the PIMS paradigm, the Austrian school, and the resource-based view of the firm. They proposed an integrative model of business performance that modeled firm-specific intangibles, industry structure, and competitive strategy variables as the major determinants of business performance. Vargo and Lusch focused attention on consumer reference scales, the psychological scales used to make evaluations of marketing-related stimuli, in consumer satisfaction/dissatisfaction (CS/D) and service quality (SQ) research and proposed social judgment-involvement (SJI) theory as a potential theoretical framework to augment, replace, and/or elaborate the disconfirmation model and latitude models associated with CS/D and SQ research. Finally, Malhotra, Charles, and Uslay reviewed the literature focusing on the methodological perspectives, issues, and applications related to correspondence analysis. They concluded with a list of the creative applications and the technique's limitations.

Articles in the Second Volume

The second volume continued the emphasis of the first by featuring a broad range of topics contributed by some of the top scholars in the discipline. The diverse articles in the second volume can all be grouped under the broad umbrella of consumer action. Bagozzi developed a detailed framework for consumer action in terms of automaticity, purposiveness, and self-regulation. MacInnis, Patrick, and Park provided a review of affective forecasting and misforecasting. Ratchford, Lee, and Talukdar reviewed the literature related to use of the Internet as a vehicle for information search. They developed and empirically tested a general model of the choice of information sources with encouraging results. Miller, Malhotra, and King reviewed the categorization literature and developed a categorization-based model of the product evaluation formation process, which assists in the prediction of set membership (i.e., evoked, inert, or inept). Lam and Parasuraman proposed an integrated framework that incorporated a more comprehensive set of various individual-level determinants of technology adoption and usage. Recently, marketing has come under increased pressure to justify its budgets and activities. Lehmann developed a metrics value chain to capture the various levels of measurement employed in this respect. Finally, Oakley, Iacobucci, and Duhachek provided an exposition of hierarchical linear modeling (HLM).

Articles in the Third Volume

Bolton and Tarasi described how companies can effectively cultivate customer relationships and develop customer portfolios that increase shareholder value. They reviewed the extensive literature on customer relationship management (CRM), customer asset management, and customer portfolio management, and summarized key findings. They examined five organizational processes necessary for effective CRM: making strategic choices that foster organizational learning; creating value for customers and the firm; managing sources of value; investing resources across functions, organizational units, and channels; and globally optimizing product and customer portfolios.

Chandrasekaran and Tellis critically reviewed research on the diffusion of new products primarily in the marketing literature and also in economics and geography. While other reviews on this topic are available, their review differs from prior ones in two important aspects. First, the prior reviews focus on the S-curve of cumulative sales of a new product, mostly covering growth. Chandrasekaran and Tellis focused on phenomena other than the S-curve, such as takeoff and slowdown. Second, while the previous reviews focus mainly on the Bass model, Chandrasekaran and Tellis also considered other models of diffusion and drivers of new product diffusion.

Eckhardt and Houston reviewed, compared, and contrasted cultural and cross-cultural psychological methods. They presented the underlying conceptions of culture that underpin both streams, and discussed various methods associated with each approach. They identified the consumer research questions best answered using each approach and discussed how each approach informs the other. Finally, they examined how consumer research can benefit from understanding the differences in the two approaches. While cultural and cross-cultural perspectives adopt distinct views about culture and psychological processes, it is possible to view them as complementary rather than incompatible. Several suggestions by Malhotra and his colleagues can be useful in this respect (Malhotra 2001; Malhotra, Agarwal, and Peterson 1996; Malhotra and Charles 2002; Malhotra and McCort 2001; Malhotra et al. 2005). For example, one can start with an etic approach and then make emic modifications to adapt to the local cultures. Alternatively, one can start with an emic perspective and then make etic adaptations to get an understanding across cultures. This systematic theory building and testing process is illustrated by Kim and Malhotra (2005).

Grewal and Compeau synthesized research from consumer behavior, psychology, and applied economics to address how price as an information cue affects consumers' responses in the context of other information cues. They developed a conceptual framework, using adaptation-level theory and transaction utility theory, that synthesized prior research on price, reference price, and other information cues and their effects on consumers' price expectations, evaluations, and behavioral intentions. Their conceptual model contributes to our understanding of the way imperfect information affects consumers' decision processes, goes well beyond the original price–perceived quality paradigm, and integrates knowledge from consumer research, psychology, and applied economics.

Sayman and Raju provided a review of research on store brands. Their review focused on integrating research in key areas and identifying directions for future research. There is limited theoretical and empirical research regarding optimal counterstrategies of national brands against store brands; studies tend to focus on one aspect, and national brand quality is typically assumed to be exogenous. Researchers have, by and large, focused on me-too-type store brands. Future research should consider premium store brand products as well.

Merunka and Peterson examined an intrapersonal aspect of language, namely, whether the structure of a language, per se, influences the thoughts of those who speak it. They reviewed empirical research conducted over the past half-century on the effects of language structure on a variety of mental activities. They found support for the weak form of the linguistic relativity hypothesis, the notion that the structure of a language does indeed influence (but not determine) cognition. The estimation of independent and joint effects of language is difficult at best. We need comprehensive studies that incorporate the order in which bilinguals acquire their respective languages, how they acquire their languages, and when they acquire their languages. Future research should also compare the possible influence of a single language on mental processing across different cultures.

Belk discussed the implications of getting visual for research, teaching, and communicating. He identified basic opportunities, threats, and consequences of becoming visual. Several techniques for collecting visual data were discussed in the realm of interviewing as well as observation. We might well be entering a Golden Age of visual and multimedia marketing research and Belk helps us to get a good handle on it.

Articles in This Volume

Consistent with the first three volumes, this fourth volume also features a broad array of topics with contributions from some of the top scholars in the field. These articles fall under the broad umbrella of the consumer and the firm.

Louviere and Meyer consider the literature on behavioral, economic, and statistical approaches to modeling consumer choice behavior. They focus on descriptive models of choice in evolving markets, where consumers are likely to have poorly developed preferences and be influenced by beliefs about future market changes. They call for a better alliance among behavioral, economic, and statistical approaches to modeling consumer choice behavior. Economic and statistical modelers can constructively learn from behavioral researchers. An understanding of the actual process that is driving preferences can provide better a priori insights into the model structures and best descriptive account of choices. The authors posit that primitive pattern-matching heuristics, which behavioral researchers suggest often underlie choices in new markets, can manifest themselves in complex functional forms of algebraic choice models, and failing to model the variance in the observed components of utility can result in misleading conclusions about the actual amount of

heterogeneity that exists in a market. They also illustrate the benefits of a reverse dialogue, how economic theory can lead behavioral researchers to more parsimonious explanations for apparent anomalies in choice tasks where preferences are uncertain.

Folkes and Matta identify factors that influence how much an individual consumes on a single usage occasion by drawing on research in consumer behavior as well as allied disciplines. They develop an integrated framework to understand how, and at what stage, various factors affect usage quantity based on Gollwitzer's (1996) "action goals" model. Initially, factors such as a product's price and social norms influence consumption-related goals and their perceived desirability and feasibility. In the next phase, factors such as self-control strategies and product instructions influence the implementation of the goal. Finally, the consumer's motivation to use feedback, and the type of feedback about consumption, has an influence on subsequent goal setting. Their framework can aid marketers in formulating products, designing packaging, and creating messages. It can also help public policy makers identify effective strategies to promote the well-being of consumers and of the environment.

Kumar and Luo also examine consumption, but from a modeling perspective. In order to allocate scarce marketing resources efficiently and effectively, it is important for a firm to know what to sell, when to sell, and to whom. Kumar and Luo review how the purchase timing, brand choice, and purchase quantity decisions have been modeled historically, as well as the issues within each decision that have been addressed. A vast majority of these studies use scanner data or transaction data. Since recent research has shown that common method variance may not be a serious problem (Malhotra, Kim, and Patil 2006), surveys can also be a useful source of such data and should be increasingly employed. They also examine the differences among various approaches and describe the common methods that have been used to model at least two of the three decisions. Finally, they describe the managerial implications of modeling these decisions and suggest ways to address the future challenges.

Despite the interest in global branding, studies involving brand extension strategies in foreign markets remain very limited. The fact that so few studies exist limits our understanding of effective brand extension strategy in a cross-cultural context. Merz, Alden, Hoyer, and Desai propose a new conceptual framework and several propositions regarding effective global brand extension strategy in a cross-cultural context. In doing so, they first review more commonly examined antecedent variables of (national) brand extension evaluation. Then, they propose a definition of culture and subsequently review the existing cross-cultural brand extension research. They examine ways in which culture may affect consumers' brand extension evaluation and develop propositions that are in need of empirical validation. These propositions are developed by drawing upon Hofstede's cultural dimensions and Roth's theory of socioeconomics and serve as a cross-cultural brand extension conceptual framework for stimulating research about brand extensions across cultures.

Given the growing importance of visual marketing in practice, Wedel and Pieters review eye-tracking research in marketing and evaluate its effectiveness. Specifically, they review eye-tracking applications in advertising (print, TV, and banner), health and nutrition warnings, branding, and choice and shelf search behaviors. They also provide a case study of the application of eye-tracking to ad pre-testing. Finally, they discuss findings, identify current gaps in our knowledge, and provide an outlook on future research.

Singh and Saatcioglu review different approaches for examining role theory implications for boundary spanners such as salespeople, frontline, and customer contact employees. They focus on universalistic and contingency approaches and develop the configural approach by extending configurational theory principles to role theory. Their effort is welcome, as neither the contingency nor the configural approach has received much attention in the marketing literature. They

compare and contrast different approaches and review literature that has remained less accessible to marketing researchers. They discuss underlying assumptions and press for critical assessment of their ecological validity. Finally, they identify promising but as yet uncharted approaches.

John considers price contract design templates governing procurement and marketing of industrial equipment. He argues that price formats choices precede the selection of a price level. These price formats are an integral aspect of the institutional arrangement devised to govern an exchange. John reviews institutions, that is, rules of interaction that govern the behavior of actors in dealing with other actors, with a focus on their pricing elements. He develops a design protocol and illustrates it by applying it to (a) the choice of fixed versus cost-plus prices for procuring components from a supplier, and (b) the choice of leasing versus selling price formats for industrial equipment.

It is hoped that collectively the chapters in this volume will substantially aid our efforts to understand, model, and make predictions about both the firm and the consumer and provide fertile areas for future research.

References

Gollwitzer, Peter M. 1996. "The Volitional Benefits of Planning." In *The Psychology of Action*, ed. Peter M. Gollwitzer and John A. Bargh. New York: Guilford Press.

Kim, Sung, and Naresh K. Malhotra. 2005. "A Longitudinal Model of Continued IS Use: An Integrative View of Four Mechanisms Underlying Post-Adoption Phenomena." *Management Science* 51 (5) (May): 741–755.

Malhotra, Naresh K. 2001. "Cross-Cultural Marketing Research in the Twenty-First Century." *International Marketing Review* 18 (3): 230–234.

Malhotra, Naresh K., James Agarwal, and Mark Peterson. 1996. "Cross-Cultural Marketing Research: Methodological Issues and Guidelines." *International Marketing Review* 13 (5): 7–43.

Malhotra, Naresh K., and Betsy Charles. 2002. "Overcoming the Attribute Prespecification Bias in International Marketing Research by Using Nonattribute Based Correspondence Analysis." *International Marketing Review* 19 (1): 65–79.

Malhotra, Naresh K., Sung Kim, and Ashutosh Patil. 2006. "Common Method Variance in IS Research: A Comparison of Alternative Approaches and a Reanalysis of Past Research," *Management Science* (December): 1865–1883.

Malhotra, Naresh K., and Daniel McCort. 2001. "A Cross-Cultural Comparison of Behavioral Intention Models: Theoretical Consideration and an Empirical Investigation." *International Marketing Review* 18 (3): 235–269.

Malhotra, Naresh K., Francis Ulgado, James Agarwal, G. Shainesh, and Lan Wu. 2005. "Dimensions of Service Quality in Developed and Developing Economies: Multi-Country Cross-Cultural Comparisons." *International Marketing Review* 22 (3): 256–278.

CONTENTS, VOLUME 1

CONTENTS, VOLUME 2

CONTENTS, VOLUME 3

REVIEW OF MARKETING RESEARCH

CHAPTER 1

FORMAL CHOICE MODELS OF INFORMAL CHOICES

What Choice Modeling Research Can (and Can't) Learn from Behavioral Theory

JORDAN J. LOUVIERE AND ROBERT J. MEYER

Abstract

In this paper we illustrate the benefits of forging a better alliance among behavioral, economic, and statistical approaches to modeling consumer choice behavior. We focus on the problems that arise when building descriptive models of choice in evolving markets, where consumers are likely to have poorly developed preferences and be influenced by beliefs about future market changes. We illustrate how understanding the actual process that is driving preferences can provide analysts with both better a priori insights into the model structures that are likely to provide the best descriptive account of choices in such settings, as well as how stable these structures are likely to be over time. We show, for example, that analogical reasoning heuristics—a common strategy for making decisions under preference uncertainty—can produce choice patterns that resemble the output of complex nonlinear, nonadditive, multi-attribute utility rules. Likewise, because novice consumers are likely to display strong individual differences in the variance of unobserved components of utility, methods that fail to recognize such differences will tend to overstate the actual extent of taste heterogeneity that exists in a population. We also illustrate the benefits of a reverse dialogue, examining how economic theory can lead behavioral researchers to more parsimonious explanations for apparent anomalies in choice tasks where preferences are uncertain. We show, for example, that some ad hoc models that have been used to statistically describe the compromise effect in choice can be deduced from first principles of rational risky decision making.

Introduction

Choice modeling research in marketing has evolved through the interplay of three different approaches to the study of human decision making. One approach is the economic perspective, which sees consumers as making choices in a manner that is consistent with random utility maximization. Consumers are viewed as having well-developed preference functions defined over product attributes, and they choose those options whose attributes offer the most attractive tradeoffs either at the time of choice or in the short or long run (e.g., McFadden 1981). A second approach is exemplified by behavioral researchers and psychologists, who argue that actual choice processes

may be far removed from the rational mechanisms assumed by economists. That is, to the extent that preferences exist at all, they are discontinuous and imprecise, with choices being the outcome of heuristic rules that are uniquely constructed in response to the external appearance of options in choice sets (e.g., Payne, Bettman, and Johnson 1993). A third and final view, a statistical approach to modeling choices, has grown rapidly since the late 1980s. Adherents of this approach claim ideological neutrality in the debate over preferences and processes. That is, choices are simply viewed as data; any model of choice is fair game as long as it passes tests of descriptive and predictive validity in a given context (e.g., Abe 1995; ter Hofstede, Kim and Wedel 2002; Rossi, Allenby and McCulloch 2005; Kamakura and Wedel 2004). Not surprisingly, the statistical paradigm is less concerned with whether any given model can be deduced from first principles of utility maximization or cognitive theory.

Although the three approaches differ philosophically and to some extent methodologically, intuition suggests that that they might usefully converge over time as our understanding of choice behavior evolves and progresses. Yet the academic reality appears to be very different. For example, behavioral researchers have tended to focus on laboratory demonstrations of failures of the assumptions of standard economic models (e.g., context invariance) and have given limited attention to developing alternative modeling paradigms that might account for these failures (for exceptions, see, e.g., Kivetz, Netzer, and Srinivasan 2004; Tversky and Simonson 1993). Adherents of the economics view of choice modeling, for their part, have frequently been dismissive of behavioral findings, arguing that lab settings exaggerate the size of errors that would be observed in real markets or that they can be captured through complex generalizations of standard models (e.g., Machina 1982). Finally, statistical modelers have done little to resolve theoretical gaps between the behavioral and economic camps. While there is much to be learned and gained from incorporating statistical advances from discrete multivariate and Bayesian statistics in choice modeling (e.g., Rossi and Allenby 2003), there is also much to be lost by adopting a purely statistical view of what is inherently a human behavioral process.

The purpose of this paper is to take a small step toward fusing these different perspectives in the analysis of choice data. We take a limited first step by exploring one dimension of this fusion, namely what empirical economic and statistical modelers can constructively learn from behavioral researchers (and visa versa) when building models of consumer choice in evolving markets—a setting where consumer preferences are likely to be highly unstable. For example, we illustrate how primitive pattern-matching heuristics, which behavioral researchers suggest often underlie choices in new markets, can manifest themselves in complex functional forms of algebraic choice models, and how failing to model the variance in the observed components of utility can lead analysts to reach misleading conclusions about the actual amount of heterogeneity that exists in a market. In our final discussion we also illustrate the benefits of a reverse flow of learning, showing how a better understanding of the rational bases of choice under uncertainty can sometimes lead behavioral researchers to simpler explanations for laboratory choice anomalies, focusing on the particular case of the compromise-effect.

Choice and Market Evolution

Let us begin with a thought experiment that illustrates the types of modeling challenges that we try to address in this paper. Consider a simple market with one (monopoly) provider of a good, such as a monopoly provider of cable or broadband services or another public utility. In this market, consumers must decide whether or not to choose the good. At some point the good in question is launched into the market, and we assume that prior to launch, information is available about the

good, its features and price(s), and the likely launch date. Thus, prior to launch some consumers in the market are aware that the service will be provided, and have reasonably complete information about its features and likely prices. Another, probably much larger, proportion of consumers is "vaguely aware" that the service will be provided, and has incomplete information about features and possible prices. Finally, a third proportion is unaware of the good or that it will be launched.

This market is thus characterized by three stylized groups of consumers, who can be viewed as being on a continuum of awareness and informedness about the good, or they can be viewed as three discrete segments. Initially, the most aware and informed are likely to be a small minority; the vaguely aware and informed, while probably a larger proportion, also are likely to be a minority; most of the market is more likely to be unaware and uninformed. Then the good is launched, and things begin to change. To the extent that the good is of interest to consumers and they are capable of buying and consuming it, which allows them to receive the associated benefits or problem solutions the good provides, we expect the proportion of consumers who are aware and informed to grow over time. Likewise, consumers who are unaware and uninformed will gradually move into the vaguely aware and informed group, and in this way the market will evolve from the "bottom up."

A marketer who wishes to model the decision of whether consumers choose the good in this market would typically begin with the tools of random utility theory. Each consumer n in the population would be assumed to associate with the new service i a utility $U_{nit} = \beta_n' X_{it} + \varepsilon_{nit}$, where $X_{it,}$ is a vector of the measured attributes of the service (e.g., price), β_n is an associated parameter vector describing the consumer's tastes for these attributes, and ε_{nit} is an unobserved component of utility. The unobserved component ε_{nit} would typically be assumed to follow an independently and identically distributed extreme value distribution (over consumers, choice alternatives, and service characteristics). If this assumption is satisfied, the individual choice probabilities can be represented by the well-known multinomial logit model

$$P_{nit} = \frac{e^{\beta_n' X_{it}}}{\sum_k e^{\beta_n' X_{kt}} + \theta_{nt}} \tag{1}$$

where θ_{nt} is the consumer's utility for unmeasured outside goods. To extend expression (1) to the study of population or market choices, analysts typically make assumptions about how tastes β_n vary over the population. For example, if β_n can be assumed to have a stationary parametric distribution, then population choice can be modeled by a random-coefficients or mixed logit model that assumes $U_{nit} = \beta' X_{it} + \eta_{nit} + \varepsilon_{nit}$, where $\eta_{nit} = (\beta_{\text{int}} - \beta)' X_{it}$ is a random variable that captures unobserved individual departures from a common strict utility $\beta' X$ (e.g., Hensher and Greene 2003).

It should be clear that while the above approach might provide a good statistical description of the association that exists between choices and service attributes at a particular point in time (or over a series of points in time) for a particular sample of people during the course of market evolution, it captures few of the behavioral features of service-choice dynamics mentioned above. For example, the model (as formulated) does not characterize how parameter heterogeneity might be associated with factors that underlie differential levels of awareness and information possessed by consumers, the provider's decisions about communications and access, and beliefs held by consumers about the market's future (e.g., the possibility of new entrants or expectations about how the technology will evolve). While analysts may acknowledge that these associations are likely

to change as a market evolves, exactly how the changes will occur or what their trajectories will typically be lies outside the purview of the analysis. Thus, it is fair to say that the overwhelming majority of these types of models are purely descriptive with little real explanatory capability.

In the sections below we try to illustrate more precisely how real behavioral processes underlying choices in markets can manifest themselves in the data that are used to estimate reduced-form statistical models, and how overlooking behavioral processes may lead analysts to erroneous conclusions about both the nature of preferences in markets and how markets will evolve over time. More specifically, we explore what behavioral theory would predict about the empirical appearance and stability of typically estimated random utility models when they are used to describe the choice behavior of consumers who:

1. Have high levels of uncertainty about their preferences for goods in a market (both attribute valuations and weights)
2. Use heuristic short-cuts that do not utilize all the product-attribute information available to them at the time of choice
3. Have strategic foresight—that is, consider how the current choice will affect the utility gained from future choices

Modeling Choice by Naïve Consumers

One aspect of random utility theory that elicits few quarrels is the assumption that consumers are guided by a desire to choose the option that will give them the most utility or pleasure. But for many consumers, particularly those in newly evolving markets, the ability to achieve this maximization goal is inhibited by the simple fact that preferences are uncertain. For example, if a novice consumer were forced to decide whether it was worth $10 a month to adopt a broadband service that would increase download speeds by 100 kilobytes, the axioms of utility theory would not help her much to make this decision. In order to make a utility-maximizing decision, she would need to know what a kilobyte is, the amount of additional pleasure that she could expect from a 100kb increase, and how to exchange the extra pleasure for dollars—knowledge few novice consumers are likely to have. The choices we observe in new markets, therefore, reflect an ambiguous mix of enduring preferences and the heuristics consumers use to overcome the *lack* of preferences.

What are the heuristics consumers use to overcome a lack of attribute preference knowledge? The consensus view is that naïve choices are often made using analogical reasoning. That is, when a new product is encountered, consumers judge it by recalling products that they consumed in the past with similar attributes (e.g., Bradlow, Hu, and Ho 2004; Gregan-Paxton and Roedder John 1997; Norman 1988). For example, this sort of pattern-matching process is thought to explain how people with little skill in mathematics can learn to play complex equilibria in games. Instead of encoding and solving optimal strategies, most games allow players to discover equilibria simply by being willing to repeat the moves that yielded the highest payoffs in the past (e.g., Camerer and Ho 1999; Fudenberg and Levine 1998). Similarly, the expertise of wine connoisseurs probably lies less in their skills at using algebraic rules to predict quality and more in their possession of a rich memory bank of past referent examples that form the basis of evaluations (known as smell and taste memory).

The pervasive use of pattern-matching heuristics in novel product judgments was illustrated in work by Meyer (1987), who examined the process by which consumers learn to make multiattribute judgments in a novel product category. In his experiments, participants were shown a series of product profiles described by several unfamiliar attributes and levels (copper alloys generated

by different production methods), and they were asked to predict each product's likely quality (strength). Once the subjects made a prediction, they received feedback about the "true" quality of an option. Consistent with behaviors previously observed in tasks like this (e.g., Mellers 1981), after several rounds of feedback participants became quite good at making forecasts. That is, they acted "as if" they had learned the multi-attribute rule that determined true quality, and were using it to make forecasts. Yet, the surprising empirical outcome associated with this experiment was that in the course of making judgments, a subset of respondents asked to provide concurrent verbal protocols gave no indication that they actually made judgments using a multi-attribute rule. Instead, they appeared to make their judgments using a pattern-matching process whereby they made forecasts about how similar any given new profile was to a previously seen profile with known qualities. Their judgments improved over time not because they were developing better knowledge of a rule, but because the database of referent examples improved, and this data base allowed them to effectively mimic the outcomes of such a rule.

Should choice modelers be worried by this result? Not necessarily; in the same way that a game theorist would be indifferent to whether people play equilibria because they actually calculate the optimal strategy or stumble their way there by trial and error (Fudenberg and Levine 1998), one presumes that random utility theorists would be happy to view multi-attribute utility theory as an "as if" model of the way in which people evaluate options. If pattern-matching heuristics produce judgment data that are well-approximated by stable linear-additive models, then one clearly can build a paradigm around this; mathematical convenience in this case would trump behavioral realism. But *is* there such an isomorphism, and how widespread might it be?

It is easy to show that a mathematical equivalence exists between pattern-matching and linear-additive rules, but only under a limiting condition of product-class experience: when the underlying (or true) reward structure is linear-additive in attributes and decision makers have had direct experience with all of the product profiles under study in a given multi-attribute space. In contrast, the more limited a person's judgmental experience in any given context, the less linear-additivity will describe their judgments, even if the underlying reward function is linear-additive.

To demonstrate, consider the following pattern-matching model of expected valuations:

$$EV_i = \delta_{iz}(V_z) + (1 - \delta_{iz})p \tag{2}$$

where EV_i is the anticipated utility of some multi-attribute profile i, δ_{iz} is a zero to 1 bounded measure of the subjective similarity between profile i and the experienced profile z that is most similar to i, V_z is the experienced utility of profile z, and p is a judgmental prior.[1] Suppose, also for the sake of simplicity, that *experienced* utility is given by the linear-additive multi-attribute model

$$V_k = \beta' X_k \tag{3}$$

where X_k is a vector of measures of the values attributes of profile k that is linear in V, and β is an associated weight vector.

It should be obvious that expression (2) is equivalent to (3) when δ_{iz} equals 1 for all comparisons i,z—that is, when every profile has been experienced by the consumer. In this case it truly doesn't matter whether consumers literally calculate utilities to form projective judgments (expression 2) or merely act *as if* they do (expression 1). But what about a more realistic case where consumers have limited experiences in a market—that is, for any given possible complete factorial array of profiles, what happens if only a small subset have valuations? In such cases (2) and (3) will not

be equivalent; instead, consumers will reveal an "as if" multi-attribute judgment rule that departs from the asymptotic (or full-information) rule (3) in predictable ways. Specifically, if we regress projected preferences (EV_i) against the attributes of each option, we do not recover a true or asymptotic utility function (2), but instead a multi-attribute function whose form is distorted by the similarity of each i to the actually experienced profiles z, their true valuations, and the judgmental prior—that is, $EV_i = \delta_{iz}(\beta' X_z) + (1 - \delta_{iz})p$.

To illustrate the properties of such an approximation, consider the case of a simple judgment context in which consumers evaluate the attractiveness of each of a number of two-attribute options on a subjective scale. Each option is described by a pair of levels that represent a combination of two six-level attributes. Respondents judge all thirty-six combinations represented by the factorial array. As above, for simplicity we assume that the true utility that would be observed by a consumer consuming each of the profiles is given by the additive rule $V_i = x_{i1} + x_{i2}$. Upon entering the task, however, a consumer has only limited experience in the category, and has directly experienced only a small subset of the thirty-six profiles. We consider the implications of approximating a pattern-matching judgment rule (expression 1) with a linear model in two illustrative cases: (1) the consumer's previous experiences correspond to the two extremes of the attribute space (the 1,1 and 6,6 profiles, respectively), and (2) the consumer has experienced three mid-valued options: the 2,2, 3,3, and 4,4 options. To generate a numeric example we used a normalized similarity metric

$$\delta_{iz} = \underset{k}{MAX}[1-(\sum_z | x_{iz} - x_{kz} | / \sum_z | x_{\max(z)} - x_{\min(z)} |]^{\lambda}$$

where $\lambda = 4$, and assumed a prior (p) equal to the 1–12 response scale mean (6.5).

In Figures 1.1a through 1.1c we plot the resulting two-factor interaction graphs for each of

Figure 1.1a **True X1*X2 Interaction**

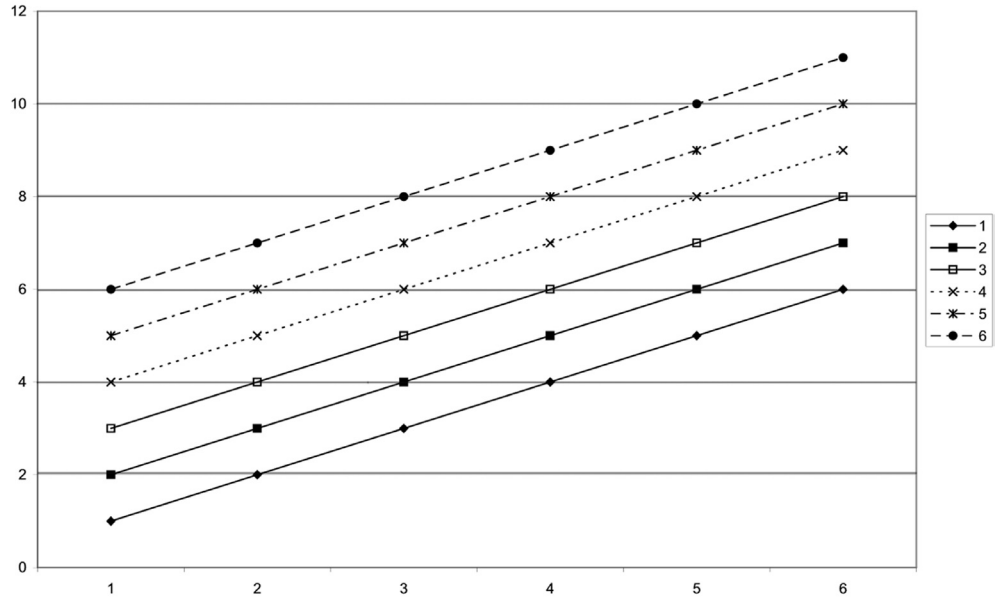


Figure 1.1b **X1*X2 Interaction when judgments are made by a pattern-matching rule with referents at 1,1 and 6,6**

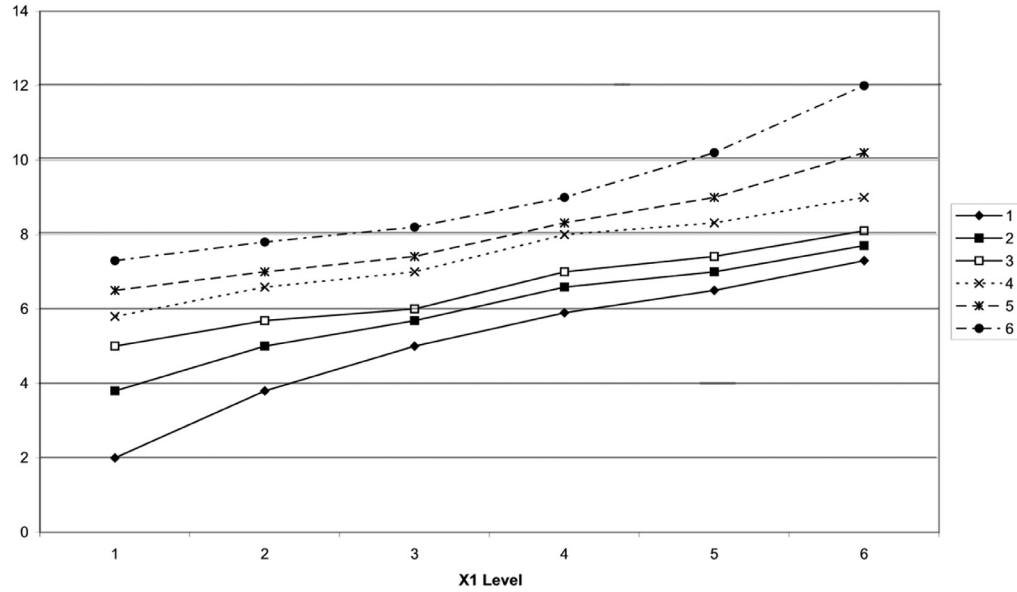
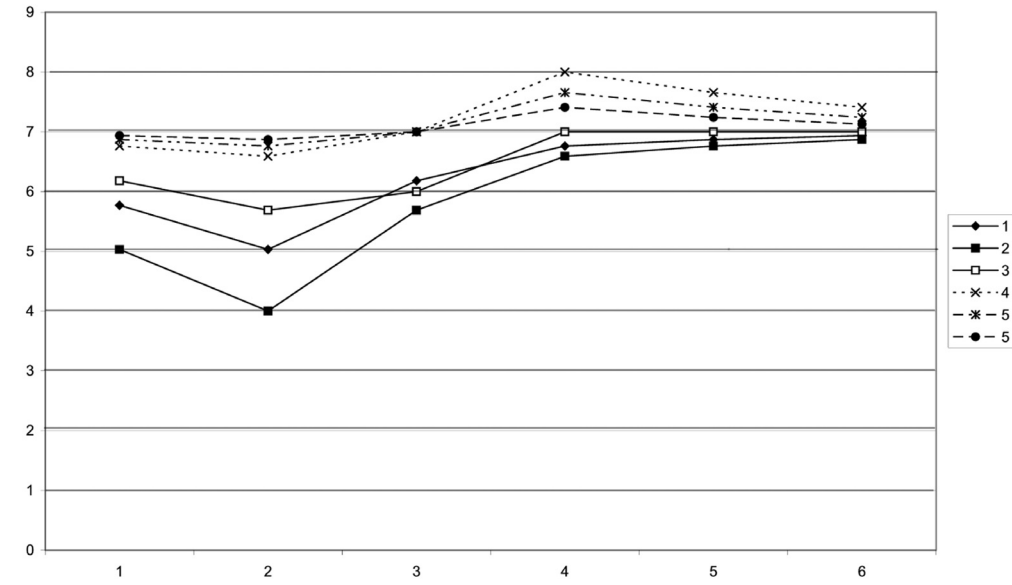


Figure 1.1c **X1*X2 Interaction when judgments are made by a pattern-matching rule with referents at 2,2, 3,3, and 4,4**



these two cases (1.1b and 1.1c) to be contrasted with the normative interaction (1.1a). Figures 1.1b and 1.1c provide good and bad news about the ability of linear models to mimic pattern-matching judgment processes. The good news is that they show that as long as a consumer's previous experiences are well chosen (in this case at the extremes of utility continua), and under an appropriate (here, neutral) prior, simple linear models can do a good job of describing contemporaneous preferences *and* yield parameter estimates that are asymptotically stable.[2] Specifically, in Figure 1.1b we see that judgments generated by a pattern-matching process that is informed by the utility extremes will forecast the true valuations of unfamiliar intermediate profiles well, and should be well described by a linear-additive model.

The bad news is that if experiences (and/or priors) are *not* well chosen, the value of linear-models will be greatly reduced in stability and descriptive validity. As shown in Figure 1.1c, when referent experiences lie in the interior of the attribute space (which may be more typical in practice), the revealed preference surface becomes nonlinear, displaying an interaction at the more distant (relative to experience) tail. This implies that not only would a simple linear model do a poor job of capturing contemporaneous preferences (one needs a data design that can estimate nonlinear effects and interactions), but it also would poorly forecast the preference structure observed at future times when the consumer's scope of experience in the category expands.

This latter result holds two implications for applied choice analysts. The first is that linear-additive models will often be ill suited for describing the association that exists between product attributes and product choices for novice consumers. Differential knowledge of utility over any multi-attribute space will produce nonlinearities and/or interactions that such models will fail to capture. But care must be taken in recalling that these interactions would not be manifestations of enduring conditional preferences (for example, a sustained increased sensitivity to service variations given higher paid prices), but rather the transient effects of limited knowledge of preferences over the attribute space. As such, they would be expected to display little temporal stability, perhaps vanishing completely as consumers become fully knowledgeable about a product category.

The Effect of Unspecified Variability in the Unobserved Components of Utility

It might be argued, of course, that because novice consumers are likely to be heterogeneous in the kinds of product experiences they have had, individual-level departures from linear-additivity due to use of naïve pattern-matching rules could well wash out when markets are viewed in the aggregate. That is, one might be able to proceed with traditional linear-additive choice models under the assumption that transient individual differences in functional form would be captured by the variance of the unobserved component of utility in a traditional linear-additive model. If the effects of subsequent learning similarly cancel themselves out across the population, then the coefficients of the linear-additive model—though perhaps wrong at the individual level—would provide good aggregate long-term forecasts of preferences.

Is the problem solved that easily? The answer, unfortunately, is "no," for two reasons. First, because choice model estimates are perfectly confounded with the variance of the unobserved component of utility (see, e.g., Swait and Louviere 1993), changes in consumer experience that alter the structure of the unobserved component over time would also induce temporal changes in these estimates. Without a theory of what is driving the error terms, the exact nature of these temporal changes would be impossible to predict. For example, consider a choice analysis that reveals novice consumers to be statistically insensitive to variations in service quality. The confound of value (coefficient) and variance implies that the meaning

Table 1.1

Consumer Preference and Variability Types

Variability X Preference	Low price sensitivity; high quality sensitivity	High price sensitivity; low quality sensitivity
Low Precision	1. $V_i = 1.5Q_i - 0.5P_i$ scale = 0.6	2. $V_i = 0.5Q_i - 1.5P_i$ scale = 0.6
High Precision	3. $V_i = 1.5Q_i - 0.5P_i$ scale = 1.9	4. $V_i = 0.5Q_i - 1.5P_i$ scale = 1.9

Table 1.2

Estimation Realization from Table 1.1

Variability X Preference	Low price sensitivity; high quality sensitivity	High price sensitivity; low quality sensitivity
Low Variability	1. $V_i = 0.9Q_i - 0.3P_i$	2. $V_i = 0.3Q_i - 0.9P_i$
High Variability	3. $V_i = 2.85Q_i - 0.9P_i$	4. $V_i = 0.9Q_i - 2.85P_i$

of this result is fundamentally ambiguous: one could never know for sure whether consumers are truly indifferent to service or whether the true effect of service is being temporarily masked by the aggregate effect of consumers using a heterogeneous mix of pattern-matching heuristics. If the latter is the case, parameters estimated now would be of little value for long-term planning purposes.

A second, more subtle, problem is that if a population is heterogeneous in its category knowledge, the variance of the unobserved utility component should also not be constant *across* a sample at any given point in time. Hence, inferences about preference heterogeneity derived from model parameters will be confounded with knowledge heterogeneity, or variations in the standard deviation of the unobserved component of utility across consumers.

For example, consider a case in which a population is characterized by a mixture of experienced consumers who reliably choose products based on a given set of attributes, and less experienced consumers whose choices are less reliably linked to attributes (e.g., they make judgments by referring to one or two products with which they have had direct previous experience). In both groups, the consumers differ in their true sensitivity to price (that is, the sensitivity to price that would be observed if one controls for all unobserved influences on choice). So, imagine that there are four types of consumers as shown in Table 1.1: (1) low choice variability combined with low sensitivity to price and high sensitivity to quality; (2) low choice variability combined with high sensitivity to price and low sensitivity to quality; (3) high choice variability combined with low sensitivity to price and high sensitivity to quality; and (4) high choice variability combined with high sensitivity to price and low sensitivity to quality. In this table, the "scale" corresponds to the inverse variance (or precision) of the unobserved component of utility.

Now, consider parameters estimated from these four consumer types, as shown in Table 1.2. It should be noted that "scale" multiplies the systematic utility component, so the parameters in Table 1.1 are "true" parameters. Note that if we know the true parameters, we can conclude that consumer types 1 and 3 are identical and types 2 and 4 are identical, except for choice variability.

Yet, with the scale confound, we would conclude that no two of the four consumer types are alike, although we might incorrectly conclude that 1 and 4 share sensitivity to quality, while 2 and 3 share sensitivity to price. A similar result would obtain if instead of discrete consumer types/classes one had continuous distributions with the true parameters as means.

Are Variance Effects on Model Parameters Real?

An obvious objection to the above discussion and stylized example is that they only establish the *possibility* that analyses of preference heterogeneity based on standard methods may be misleading. In turn, this begs the questions as to the degree to which individuals indeed differ in the variance of the unobserved utility component, and the degree to which preference parameters are confounded by this variability.

An illustration of the magnitude and nature of variance effects was recently provided by Louviere and Eagle (2006). They report the results of 66 choice experiments whereby choice models were estimated for single individuals, which allows one to estimate the size of the variance of the unobserved component of utility in choice. To provide a flavor of these analyses, in Appendix 1.1 we report the results of 21 individual-level model estimates from 2 of the 66 experiments, reflecting choices among hypothetical pizza products and cross-country flights. These two contexts are reported simply for convenience and because they have fewer parameters than other contexts. The two experiments used a common underlying optimally efficient design to estimate the main effects of a 2^3 x 4^3 factorial based on three options per set. Participants were members of an Australian opt-in online panel; completion rates for both conditions were over 80 percent.

Tables 1A.1 and 1A.2 in the appendix display the individual-level multinomial logit model estimates for the subjects who participated in the experiments, and Tables 1A.3 and 1A.4 give summary statistics for the experiments. The individual-level model estimates allow one to calculate residuals from model predictions, in turn allowing one to regress design matrix codes on residuals. This auxiliary regression allows one to determine if (a) significant unobserved variability remains after MNL estimation, and (b) the remaining unobserved variability is systematic (i.e., residuals systematically related to design elements). Both conditions produced similar results, with 18 of the 21 individuals in the pizza condition and 17 in the flights condition exhibiting regression results that are significant at a 90 percent level. Thus, the vast majority of individuals in both conditions have significant remaining unobserved variability that is systematically related to design attribute levels. Thus, it is unlikely that the individuals satisfied constant error variance assumptions.

In Figures 1.2a and 1.2b we graph the individuals' mean squared model residuals against their airfare and price utility estimates, for flights and pizzas, respectively. Both graphs are consistent with random utility theory, which predicts that as error variances increase (measured by mean square residuals), model parameter estimates should converge to zero. Both graphs display this result, allowing one to "see" that the magnitudes of the airfare and price estimates are a function of the variability in each individual's choices. Thus, in the case of airfares and pizza prices, a large proportion of parameter differences between individuals can be explained simply by differences in individuals' choice variability. Thus, model estimates of these effects are significantly confounded with individual differences in variability.

More specific details of the breakdown of these effects are as follows:

1. Summary of variance explained by choice variability (MSR) between individuals (flights)

Figure 1.2a **MSR vs. Fare Estimates: Flights**

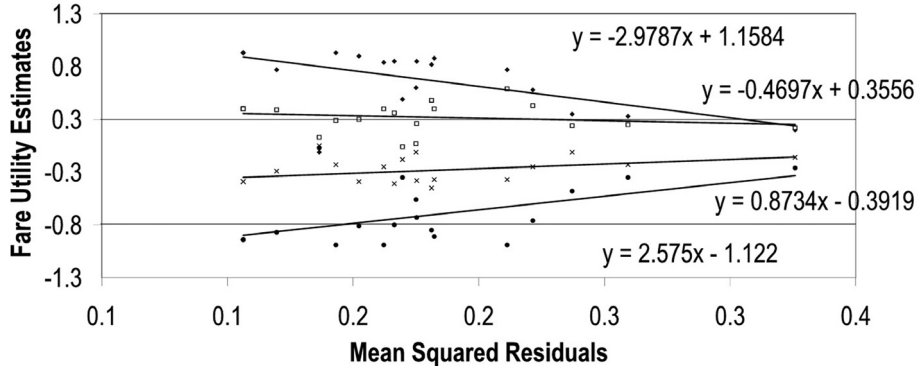
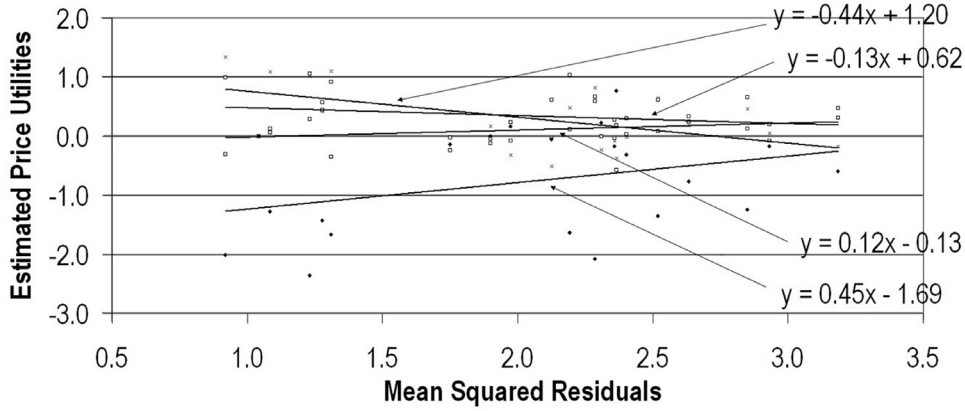


Figure 1.2b **MSR vs. MNL Price Estimates: Pizzas**



a. Travel times—6.8%
b. Airfares—22.1%
c. Airline brands—21.1%
d. Frequent flyer program—0.4%
e. Number of stops—24.5%
f. Free drinks—9.1%
g. Alternative-Specific Constants (ASCs)—20.5%
h. Average across all individual estimates—16.2%

2. Summary of variance explained by choice variability (MSR) between individuals (pizzas)

a. Pizza chain—1.2%
b. Pizza prices—21.2%
c. Number of toppings—23.6%

d. Free bread—0.0%
e. Free drinks—8.6%
f. Free dessert—12.9%
g. ASCs—19.3%
h. Average across all individual estimates—13.6%

The importance of these results is that they demonstrate the danger of interpreting empirical variance in choice model parameters as uniquely reflective of either preferences or preference heterogeneity. Specifically, if true parameters are confounded with error variances, choice models can forecast future choice behavior well only if the variances are temporally stationary and/or they do not co-vary with other factors that were constant in the data source used for estimation, but are not constant at times or places or in segments to be predicted. As noted above, there is good reason to suspect that they often will not be constant, particularly in the case of developing markets, but also in many cases where no attempts are made to understand other possible sources of variability in choices (a typical case in choice modeling).

Dealing with Complex Choices in Mature Markets: Short-Cut Heuristics and Their Representation

The above discussion focused on problems modeling choices in early stages of market evolution when the set of market alternatives is small, but the uncertainty in how to evaluate these alternatives is high. As markets approach maturity, however, the opposite problem occurs; while there may be little uncertainty in forecasting the utility that consumers will derive from individual offerings, their choices may well be more difficult if there are a large number of differentiated offerings. In this section we take up what we know about how informed consumers make choices from complex sets, and the implications of this for standard choice analyses.

A pervasive finding of work that has studied processes associated with complex choices is that decisions are often guided by noncompensatory screening rules that act either to produce unique choice outcomes or to sequentially reduce choice sets to cognitively manageable sizes (e.g., Payne, Bettman, and Johnson 1993). For example, individuals may eliminate alternatives if they (a) fail on a critical product attribute (a conjunctive elimination rule), (b) fail to offer at least one distinctive benefit across attributes (a disjunctive elimination rule), or (c) are unattractive by virtue of their rank-order position in a set (a rank elimination rule; e.g., Einhorn 1970). In theory, the use of such decision rules could be problematic for random utility models because they imply that indirect utility functions are not strictly linear and additive. Instead, options are evaluated using noncompensatory rules that make only limited use of attribute information.

As we showed earlier, under some conditions linear models can provide a reasonably close first approximation to decisions made by pattern-matching rules. Fortunately, they also can provide a reasonably close first approximation to noncompensatory screening or elimination rules—as long as one is willing to consider nonadditive forms that admit interactions among cues (see, e.g., Einhorn 1970). For example, consider a consumer who makes a series of binary judgments about whether each of several products described by two attributes, X_1, and X_2, is acceptable or not. The consumer makes her judgments using a noncompensatory conjunctive screening rule, as follows:

$$\textit{Acceptable if } [X_1 > \alpha] \textit{ and } [X_2 > \delta]. \tag{4}$$

It is easy to show that a continuous algebraic analog to (4) exists as long as there is imprecision in the attribute thresholds α and δ that drive judgments for individuals within a given data set. If α and δ act as random variables, the *probability* that a product profile (X_{i1}, X_{i2}) will be judged as acceptable is given by the product of the marginal probabilities that the realized (and unobserved) values of α and δ are less than X_{i1} and X_{i2} at the moment of choice; that is,

$$Pr(i|\, X_{i1,}X_{i2}) = Pr(Xi_1 > \alpha)\; Pr(Xi_2 > \delta). \tag{5}$$

The continuous-functional analog of (5) that follows depends on the assumed functional form of the distribution of errors associated with the acceptance thresholds α and δ. For example, if the cumulative densities of the errors associated with α and δ can be approximated by a linear probability model of the form $\Pr(X_{i1} > \alpha) = \alpha_o + \alpha_1 X_{i1}$ and $\Pr(X_{i1} > \delta) = b_b + b_1 X_{i2}$ then data generated by a conjunctive process for generating acceptability judgments would correspond to the bilinear utility function

$$\begin{aligned} Pr(i|\, X_{i1},X_{i2}) &= (a_0 + a_1 X_{i1})(b_0 + b_1 X_{i2}) \\ &= k_0 + k_1 X_{i1}; + k_2 X_{i2} + k_3 X_{i1} X_{i2} \end{aligned} \tag{6}$$

(e.g., Keeney and Raiffa 1976). Expression (6) contains an important implication, namely that if the errors are associated with thresholds and the distribution of these errors is approximately linear, conjunctive screening processes are mathematically equivalent to a linear probability model that recognizes linear-by-linear interactions between attributes. By extension, the finding of fan-like (linear-by-linear) interaction among a pair of attributes in a multi-attribute judgment experiment (e.g., a full-factorial conjoint design) has long been seen as suggesting the likely use of noisy screening rules in judgment by decision makers (e.g., Louviere 1988).

Note that we can extend this idea to any arbitrary set of conjunctive or disjunctive screening rules and error distributions. Specifically, for a set of independent acceptability judgments generated by a general family of stochastic screening rules of the form ($X >$, $<$, *or* $=\zeta$), it should be clear that there will always exist an equivalent continuous algebraic counterpart of the form:

$$\Pr(i \mid X)_n = f_n(X_i) + g_n(XX_i) \tag{7}$$

where $f_n(X_i)$ is a general polynomial expansion about the attribute vector X_i, viewed by decision maker n, and $g_n(XX_i)$ is a similar expansion over the vector of cross-products or interactions among the elements of the vector X.

The Costs of Misspecification: An Empirical Illustration

Expression (7) seems to provide a simple remedy for capturing choice behavior if one suspects that consumers use an unobserved array of noncompensatory screening rules. That is, simply construct an appropriate design that allows one to estimate a generalized set of interactions among product attributes. Such designs should allow analysts not only to capture the average effect of

using noncompensatory rules on decisions, but also heterogeneity in the structure of these rules across a population. So, it is surprising that there have been few attempts in choice-model applications to estimate such general forms. There appear to be two reasons why this is the case. The first is pragmatic, namely that ecological data like a panel are rarely rich enough to support identification of complex models like (7). The second is that laboratory choice experiments that allow such estimation impose significant data requirements that involve obtaining observations from (potentially) large factorial arrays, which historically has been seen as impractical in most field settings (although, as noted by Louviere, Hensher, and Swait, 2000, this in fact is not true).

However, a more likely reason why more complex indirect utility functions are not more commonly estimated is the long-standing result by Dawes and Corrigan (1974), namely that estimating higher-order interactions adds little to model fit or out-of-sample predictive ability. Specifically, as long as attributes are monotonic in their effect on a criterion, and attributes across alternatives in choice are not maximally negatively correlated (i.e., form a perfectly efficient Pareto set), it will be the case that a strictly linear-compensatory choice model will mimic many noncompensatory choice rules (e.g., Dawes and Corrigan 1974; Einhorn, Kleinmuntz, and Kleinmuntz 1979; Johnson, Meyer, and Ghose 1988). In short, if one only cares about statistical description and prediction, simple linear models will often be good enough for what they are commonly used for.

But what if analysts are interested in more than prediction, and want to use choice models to derive insights about processes and/or the substantive nature of preferences in a population? Now robustness no longer applies, and omitting interactions from the indirect utility function not only can lead to biased estimates, but also can lead analysts to misleading conclusions about how product attributes influence market choices.

As an example, consider what happens if one designs a typical conjoint experiment, but instead of asking individuals to rate or rank the experimental product profiles, one asks them to evaluate each option and state if they would (yes) or would not (no) choose each. To make the example concrete, we consider pizza delivery services described by four attributes (price, brand name, number of toppings, and type of crust); each attribute has two levels, and each individual is asked to evaluate and respond to the entire factorial (2^4).

We constructed fifteen hypothetical individuals, each of whom is represented by a particular deterministic decision rule to say "yes" or "no" to each pizza profile. For example, an individual might use the rule "say yes if price is low and crust is thin," or "say yes if brand is Domino's, crust is thick, and number of toppings equals 4," or "say yes if price is low and brand is Pizza Hut." We apply the 15 rules to generate the yesses and noes associated with each of the 16 experimental profiles. Thus, the dataset produced by this process contains 16 scenarios x 15 individuals, or 240 observations. In the interests of space we omit typical preliminary analyses that one should conduct on the data set, such as calculating marginal frequencies (conditional means) for each attribute level. We can summarize these analyses by noting that like almost all choice experiment data sets, the marginal frequency counts indicate that all effects are large and have acceptable signs (preference directions).

To begin our analysis of these data, we first estimate a simple, one-size-fits-all binary logit model. Model estimates and associated statistics are shown in Table 1.3. By and large, all the effects are significant, although crust type is marginally significant. Instead of standard log-likelihood results, consider how well the estimated model predicts observed response probabilities. That is, each hypothetical individual faces the same 16 scenarios; hence, we can calculate the observed proportion of yesses for each scenario, and compare this with the predicted proportion of yesses from the estimated model, allowing calculation of conventional r-square values. The simple model fits the estimation data fairly well, with an r-square value of 0.73. If presented with such a set of results, many analysts would likely conclude that their work was done; the model

Table 1.3

Effect	Estimate	StdErr	Wald	P(wald)
Pizzaname	–0.4911	0.1887	6.7697	0.0093
Pizzaprice	–0.5585	0.1916	8.4994	0.0036
Crusttype	0.2933	0.1832	2.5631	0.1094
Ntoppings	0.3588	0.1847	3.7758	0.0520
Constant	–1.8146	0.2067	77.0804	0.0000

fits well, and yields intuitively reasonable insights about how attributes affect choice (pending out-of-sample validation).

But suppose we estimate an auxiliary regression using the residuals as the dependent variable and the design matrix as covariates. Given the nature of the data-generating process, it is not surprising that we obtain a highly significant regression result ($F = 3.5$, $P(F) < 0.000$), with each main effect significant at the 90 percent level, and at least one interaction (price x number of toppings) also significant. So, it should be clear that something is wrong with the binary logit model.

To address this, we add all the two-way interactions to the one-size-fits-all binary logit, and reestimate. This model fits the estimation data significantly better, although the price x number of toppings interaction is not significant. But there is still a problem, namely an auxiliary regression analysis of the residuals from this model again produces a significant result (F=2.5, $P(F) < 0.002$); the main effects again are significant, and the price by number of toppings interaction again is significant. At this point, if we were presenting this finding to a conference, we would expect to hear the typical refrain of "you need to take preference heterogeneity into account." Of course, in the present case, it is not preference heterogeneity that is the source of misspecification, but *rule heterogeneity*; we have overlooked heterogeneity in the array of interactions that represent different noncompensatory rules.

As discussed above, each decision rule that was constructed can be represented as a Linear Probability Model (LPM). Each individual has a different LPM, and the LPMs generally will contain interaction terms. Table 1.4 below displays each of the fifteen LPMs representing the decision rules. Most rules (individuals) contain one or more zero estimates, indicating that these particular effects are not part of the rule. The table is divided into two parts: (a) the left-hand side contains estimated main effects, with r-squares for each individual in column 7; and (b) the right-hand side contains all two-way interactions, with associated r-squares for all main effects and two-way interactions in the last column. It should be obvious that all r-squares increase substantially when we add interactions. Individuals with an r-square value of 1.0 are fully described by a rule that requires only main effects and two-way interactions; individuals with r-square values of less than 1.0 require additional interactions that we omit to save space.

This analysis gives us a much richer and more accurate view of what drives choices in this market. For example, it highlights that various product attributes affect choice not as independent main effects but rather as interactions with other attributes, and also that the pattern of these interactions varies considerably within the population.

Other Approaches to Capturing Rule Heterogeneity

It is important to emphasize that this is but one of a number of approaches that have been suggested over the years for representing individual noncompensatory choice process in choice models, each having its own comparative strengths and weaknesses (e.g., Gilbride and Allenby 2004). For ex-

Table 1.4

LPM Estimates for 15 Rules

Ind	Const	Brand	Price	Crust	No. of tops	Main Effects R-Sq	Brand price	Brand crust	Brand ntops	Price crust	Price ntops	Crust ntops	Main Eff+Int R-Sq
1	0.13	–0.13	–0.13	–0.13	0.00	0.43	0.13	0.13	0.00	0.13	0.00	0.00	0.86
2	0.25	–0.25	0.00	0.00	–0.25	0.67	0.00	0.00	0.25	0.00	0.00	0.00	1.00
3	0.25	–0.25	–0.25	0.00	0.00	0.67	0.25	0.00	0.00	0.00	0.00	0.00	1.00
4	0.06	0.06	–0.06	0.06	0.06	0.27	–0.06	0.06	0.06	–0.06	–0.06	0.06	0.67
5	0.06	–0.06	–0.06	0.06	0.06	0.27	0.06	–0.06	–0.06	–0.06	–0.06	0.06	0.67
6	0.13	0.00	–0.13	–0.13	0.13	0.43	0.00	0.00	0.00	0.13	–0.13	–0.13	0.86
7	0.25	0.25	0.00	0.25	0.00	0.67	0.00	0.25	0.00	0.00	0.00	0.00	1.00
8	0.13	–0.13	0.00	–0.13	0.13	0.43	0.00	0.13	–0.13	0.00	0.00	–0.13	0.86
9	0.25	–0.25	0.00	0.25	0.00	0.67	0.00	–0.25	0.00	0.00	0.00	0.00	1.00
10	0.13	0.00	–0.13	0.13	0.13	0.43	0.00	0.00	0.00	–0.13	–0.13	0.13	0.86
11	0.25	0.00	–0.25	0.00	0.25	0.67	0.00	0.00	0.00	0.00	–0.25	0.00	1.00
12	0.25	0.00	–0.13	0.00	0.00	0.08	–0.13	0.00	0.25	–0.13	–0.13	0.00	0.67
13	0.25	0.00	0.25	0.25	0.00	0.67	0.00	0.00	0.00	0.25	0.00	0.00	1.00
14	0.13	–0.13	–0.13	0.00	0.13	0.43	0.13	0.00	–0.13	0.00	–0.13	0.00	0.86
15	0.06	–0.06	–0.06	–0.06	0.06	0.27	0.06	0.06	–0.06	0.06	–0.06	–0.06	0.67

ample, a major downside of representing noncompensatory heuristics using generalized families of interactions as illustrated above is that interaction parameters do not have a direct translation to specific noncompensatory heuristics. Moreover, interpretability of parameter estimates is further confounded by the existence of other more mundane sources of interaction effects on choices, such as levels of quality being interpreted differently depending on the observed price.

Alternative proposals for capturing rule heterogeneity that could avoid these ambiguities have been provided by (inter alia) Elrod, Johnson, and White (2005), Gilbride and Allenby (2004), and Swait and Adamowicz (2001), who describe generalized choice models that recognize the existence of a mix of compensatory and noncompensatory choice heuristics. However, a limitation of these proposed approaches is that they capture variation in only a small set of pre-specified heuristics (e.g., compensatory versus conjunctive screening rules). If consumers make decisions using a mix of rules other than those assumed by these models (likely to be the case), their value as descriptive (and possibly predictive) tools would likely be less obvious.

The Problem of Consumer Foresight

The final topic we take up is a challenge to choice modelers in all phases of market evolution, namely that consumers make choices with knowledge of and expectations about future consequences. An often-cited limitation of standard choice analyses is that they assume consumers are not only utility maximizers, but *myopic* utility maximizers, whereby their goal is to choose that option that offers the highest expected pleasure at the time of choice, without thinking about how this choice might affect the utility/pleasure associated with future choices (e.g., Erdem and Keane 1996). Thus, few standard analyses allow for the possibility that a consumer might choose an option for the mere purpose of gathering information about it, delay choice out of a belief that a better choice set will be available at a later point in time, or elect not to choose an obviously superior option in order to savor the anticipation of its later consumption. Naturally, standard models have no problem *describing* choices in such settings; a decision to delay consumption, for example, can be well described by a model that posits low utility for the good at the time of choice. But such a mere statistical description clearly is dissatisfying because it ignores dynamics that produce the behavior (the distinction between not choosing and delaying choosing), and such a model only can predict behavior well in contexts identical to that in which it was estimated.

A perhaps more serious consequence of overlooking dynamics is that reduced-form or cross-sectional models of dynamic processes often will yield parameter estimates that suggest that consumer decisions are less rational than they really are. For example, a classic finding involves upward-sloping contemporaneous demand curves. If consumers believe that prices set by sellers will be higher tomorrow than they are today, short-term price increases may display positive elasticities (see, e.g., Erdem, Imai, and Keane 2003). The reason is not that consumers prefer higher prices to lower prices, but instead that their buying decisions are made in light of their beliefs about what *future* prices will be, which can give the appearance of a positive short-term reaction to observed price increases.

Due to increased recognition of these concerns, a major growth area in choice modeling research in marketing has been dynamic structural models that explicitly assume that consumers are multi- rather than single-period choice optimizers (e.g., Erdem and Keene 1996; Gonul and Srinivasan 1996; Erdem, Imai, and Keane 2003; Song and Chintagunta 2003). General acceptance of this work, however, has been limited by two factors. One factor is the pragmatic problem of computational complexity, such that efficient ways to empirically solve complex dynamic optimization problems have become available only recently, and are not part of the standard set of estimation

tools familiar or available to applied choice modelers. This limitation may only be temporary, but the second factor is more basic, namely, the complexity of such models is viewed by some as an unreasonable behavioral description of consumer planning (Houser, Keane, and McCabe 2004). Specifically, given the well-known finding that individuals find it difficult to make once-off decisions in an optimal manner (e.g., choosing gambles in the way prescribed by expected utility theory), intuition suggests that they would have little chance to optimally solve more complex dynamic programming problems. Yet, this is exactly what dynamic structural models assume that consumers are able to do.

But is this intuition correct? It is important to remember that in economics the acid test of whether a given model is theoretically tenable is *not* whether it is cognitively realistic (probably few are), but instead whether it describes equilibrium behavior that could be reached by evolutionary processes. That is, boundedly rational decision makers need only be more prone to repeat actions that tend to give higher payoffs. The fact that consumers make little attempt to plan ahead or have no idea what "backward induction" means does not preclude them from acting *as if* they do.

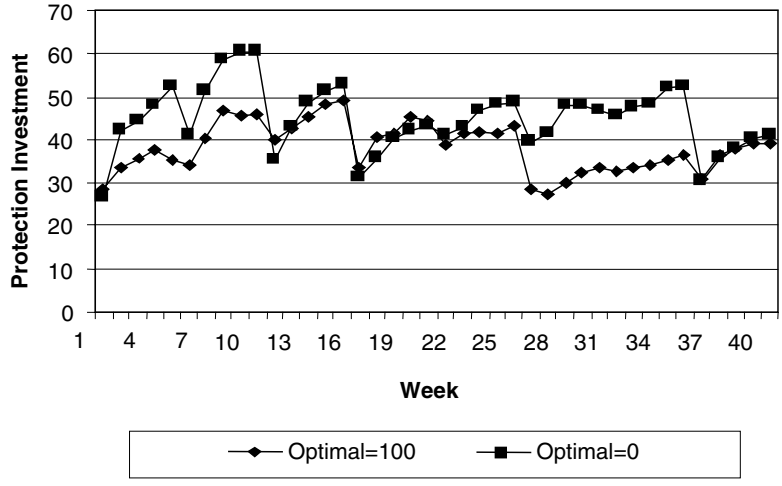
As an example of this, Hutchinson and Meyer (2005) recently reported the results of a study examining the ability of people to make accurate judgments about the expected maxima of distributions, an ability assumed in most optimal dynamic decision models (e.g., job search models). They studied this in two related experiments. The first was a paper-and-pencil task in which participants were asked to provide intuitive estimates of the expected value of the largest number that would be realized from N draws from a 0–100 uniform distribution, where N varied from 2 to 10. Perhaps not surprisingly, participants did not perform this task well, displaying a consistent tendency to underestimate the true maxima. For example, for the case of two draws (the simplest case), where the normative answer is 67, the mean estimate was 58 with individual guesses ranging from 50 to 65. Given this result, one might think it reasonable to question any model that assumes that people are good intuitive judges of maximum order statistics. But would that be correct?

To answer this, a second experiment was conducted in which participants were asked to play a lottery game for money that required implicit rather than explicit knowledge of expected maxima. Specifically, subjects played a computer game that required them to place a bet on which of two sets of N draws from a uniform urn would yield the higher maximum. In the game, participants first observed a simulated dealer take N draws from the urn (N varied across tasks). After seeing the maximum value drawn by the dealer, they were then asked to indicate whether this value was likely to be higher than that which would be realized given N new draws. After participants made this prediction, N new draws were taken and the outcome revealed. If their directional bet proved correct they received a small cash payment, but they received nothing if their bet proved wrong. Participants placed 30 such bets within each of 3 draw-size conditions (2, 4, and 6).

While participants might not be able to compute the maxima of distributions, it turns out that they can play *as if* they can. Across all three N-size conditions and trials, participants placed the normatively correct bet more than 80 percent of the time. Moreover, they acquired this "as if" knowledge quite rapidly. For example, while success rates on the first trial averaged slightly more than 50 percent (consistent with random guessing), by the second trial it reached 75 percent, and on the third 82 percent. Hence, much like the famous example of pool players and physics, the requirement for behavior to be optimal is not that people can compute optima, but rather that they live in a world that naturally reinforces optima.

But such optimism about "as if" optima comes with a strong word of caution, namely that the world does not provide consumers with frequent opportunities to learn, or reliably reward, optimal behavior. For example, in many cases the optimality of strategic choice policies cannot be observed in the short run, but only by their repeated application over a long-time horizon (such as

Figure 1.3 **Optimal versus Actual Protection Levels Over Time**



in the case of dieting). The more short-sighted a consumer, or the more stochastic the short-term reward, the more she is likely to stray from optimality by basing decisions on tangible short-term consequences rather than less tangible long-term ones.

To illustrate this, Meyer and Kunreuther (2005) describe the results of a computer simulation designed to assess the ability of people to make optimal decisions to invest in long-term protection against a low-probability, high-consequence hazard, which in this case is an earthquake. In the simulation, participants were asked to imagine that they would be living for five years in a country that was prone to periodic earthquakes. They were given a home with a certain fixed value, and as time elapsed in the simulation they had the option to invest some of this home value in permanent home improvements that would potentially lower the home's vulnerability to earthquake damage, or instead to invest it in a bank at a fixed interest rate. At the end of the simulation they were paid an amount tied to original home value minus losses due to earthquake damage and investments in protection.

The central manipulation was whether or not these investments in protection were truly effective. Participants were told that there was a 50 percent chance that the investments did little to reduce damage, and a 50 percent chance that they would be highly effective. To allow learning, they played eight rounds of the simulation, during which they had an opportunity to observe not only the damage that their own home suffered from earthquakes conditional on their investments, but also the investments and damage made associated with other players.

In Figure 1.3 we graph the actual percentage of investments over each five-year block of the simulation conditional on the optimal level of mitigation (conditional on knowing its true effectiveness). The figure shows a disturbing result, namely that not only were the participants unable to learn the investment optimum, but the mean investment level was systematically *higher* when investments did nothing to lower earthquake losses.

What explains this dysfunctional pattern of behavior? The answer is twofold. First, a statistical analysis of the period-by-period investment decisions indicated that the primary driver of decisions to invest was the magnitude of loss actually experienced in the previous period. Hence, effective mitigation had a paradoxically *negative* effect on a participant's perceived urgency to invest more when it was really effective. The second reason is that because quakes were infrequent, the ben-

efits of protection were rarely immediately seen. Instead, it was more common for participants who invested in protection to find out that it was not immediately needed than to find out it was immediately useful, a factor that further suppressed interests in buying.

The contrasting nature of these two examples underscores the fact that the choices we observe in markets are much more a reflection of the structure of the environmental feedback received by consumers rather than intuitive math abilities. The optimistic implication of this for applications of dynamic structural models is that the criticism of their validity as a literal account of how decisions are made is largely irrelevant to their status as useful empirical theories. More precisely, the fact that the component mathematical assumptions of a model (e.g., that people can accurately judge order statistics) is empirically invalid does not imply empirical failure of the holistic predictions of an associated model. Rather, that outcome depends on something different, namely whether consumers can observe feedback from the choices they make in markets, and whether the nature of this feedback in the long run will favor economically optimal behaviors over suboptimal ones.

Discussion

This paper was motivated by a desire to enhance cross-learning among the three traditionally disparate behavioral, economic, and statistical approaches to studying individual choice behavior. Central to the discussion was the suggestion that at best most current choice models offer a static snapshot of preferences in markets. The estimated parameters of choice models reflect not just the enduring preferences that consumers may have for products and attributes, but also the momentary state of consumer learning about products and/or heuristic shortcuts they use to deal with the complexity of markets. That is, the models provide something akin to a one-dimensional view of a multidimensional process. In our view, the key to building better planning models is not to build ever-more complex and accurate statistical descriptions of such a one-dimensional projection, but instead to try to better understand the multidimensional process that underlies the projection.

Although this paper focused on how behavioral theory can help choice modelers gain better understanding, it is important to emphasize that a reverse flow can be no less valuable. That is, economic theory can be useful in allowing behavioral researchers to build better models for explaining the simple mechanics that actually underlies choice anomalies or laboratory findings where behavior appears to depart from the prescriptions of rational choice theory.

Reversing the Dialogue: An Example

To illustrate, consider multi-attribute choice models developed to account for an anomaly in choice behavior known as the *compromise effect* (Tversky and Simonson 1993). A simplified account of the effect is that when participants in choice experiments are shown an array of options arrayed along a Pareto frontier in a multi-attribute space, there is a tendency for the aggregate choice portions to be massed toward the center, regardless of where the choice options are aligned along the frontier. Hence, one can increase the odds of an option being chosen simply by framing it as the compromise alternative in a set.

This effect is termed an anomaly because, taken at face value, it violates the fundamental property of random utility theory known as regularity. That is, one should not be able to increase the odds of choosing an alternative in a choice set by enlarging the set (i.e., the choice probabilities should obey regularity). Yet, studies of the compromise effect suggest that just such an effect is

possible, namely that one can increase the choice share of an extreme option by introducing a new option that is even more extreme.

Several researchers have observed that this effect can be reconciled within a random-utility framework by assuming that consumers make choices using strict utility functions where attribute values are assessed vis-à-vis choice set-specific extremes (e.g., Kivetz, Netzer, and Srinivasan, 2004; Sheng, Parker, and Nakamoto 2005; Tversky and Simonson 1993). For example, Kivetz, Netzer, and Srinivasan (2004) showed that compromise-effect data are well fit by a "contextual concavity model":

$$U_{nu} = b_i + \sum_k b_k (x_{ik} - x^S{}_{\min,k})^{c_k} \tag{8}$$

where X^S min, k is the smallest observed value of attribute k within choice set S and c_k is an empirical shape or concavity parameter for attribute k (Kivetz , Netzer, and Srinivasan 2004; Sheng, Parker, and Nakamoto 2005 offer similar forms).

Does expression (8) provide a useful theoretical account of the compromise effect? While there is ample evidence that it offers a good *statistical description* of the effect, its value as a *theoretical explanation* is less obvious. The central issue is that the compromise effect is not a universal phenomenon, but rather is commonly observed only under restricted laboratory conditions where:

1. Participants are uncertain how to value and trade off attributes (the effect does not work, for example, for choices among mixtures of monetary payoffs)
2. Choices are made by different groups of subjects viewing different choice sets with no feedback

Preference uncertainty, however, is not explicitly modeled in (8), making it an incomplete account of the phenomenon. While the model can statistically describe compromise effects in laboratory tasks designed to create it, it cannot endogenously predict what would happen if we were to alter some of the basic conditions of the task, such as reducing uncertainty through learning.

How might one build a model of the task where uncertainty is endogenous? One possible—though unlikely—starting point is to assume that participants deal with preference uncertainty in the way that would be prescribed by rational theories of risky choice. It turns out that doing so leads to a surprising result: not only we are led to a model that endogenously recognizes preference uncertainty, but the model also reveals that the compromise effect may not be an "anomaly" at all.

To see this, imagine that you are invited to play a gamble in which you are offered three options, each described by a value on two attributes that are expressed in arbitrary units of measurement called "ps" units and "kz" units:

	Option		
	A	**B**	**C**
Attribute 1	75ps	50ps	25ps
Attribute 2	2kz	5kz	8kz

Each of the units of measurement has a linear rate of conversion to a dollar payoff, but the nature of this conversion is unknown. Specifically, you only know that for each attribute *i* there is a payoff $P_i = a_i + b_iX_i$, where P_i is the payoff in dollars, X_i is the observed value of *i* (expressed in units of *ps* or *kz*), and a_i and b_i are realizations of random variables with joint distribution $f(a_i,b_i)$, such that $b_i \geq 0$. Your goal is choose the option that delivers the highest joint payoff across both attributes.

Although highly abstract, the task should be recognized as capturing the essential uncertainty that participants face in compromise experiments. One is asked to make a choice among options in which one is unsure about the mapping that exists between attribute values and utility (other than more is better), and where one has no opportunity to learn these tradeoffs by choice experience.

Is there a rational solution to this problem? There is, and it is actually quite simple. First, because the original units of measurement are arbitrary, and constant scale differences *between* attributes do not affect the solution to the choice problem, we can reduce the dimensionality of the joint distributions $f(a_i,b_i)$ by rewriting the matrix above in an equivalent normalized form, that is:

	Option		
	A	**B**	**C**
Rank (Attribute 1)	1	.5	0
Rank (Attribute 2)	0	.5	1

If $g(b_i)$ is the resulting marginal distribution of b_i, each option thus has expected payoff

$$E(P_i) = v(Z_{i1}\int_{b_1} b_1 g(b_1)db_1) + v(Z_{i2}\int_{b_2} b_2 g(b_2)db_2) \tag{9}$$

$$= v(Z_{i1}\overline{b_1}) + v(Z_{i2}\overline{b_2}) \tag{10}$$

where Z_{ij} is the normalized score of option *i* on attribute *j*, and $v(\cdot)$ is the decision maker's marginal value function for money.

Expression (9) makes a simple (and quite intuitive) prediction: for the current example were $Z_{i1} + Z_{i2} = 1$, under uniform priors (i.e., $\overline{b_1} = \overline{b_2}$), a risk-neutral decision maker for whom $v(\cdot)$ is linear would be indifferent among the options. That is, he or she would recognize that there is no one best answer to the problem as long as the attribute-payoff conversions are unknown. On the other hand, if the decision maker were *risk averse*—that is, the person has a value function that is strictly concave over $Z\bar{b}$—then she should pick the middle or compromise option.[3]

Now, here is the critical step. In the typical experimental setup associated with demonstrations of the compromise effect (see, e.g., Kivetz, Netzer, and Srinivasan 2004), an experimenter presents *different but overlapping* choice sets to *two different groups* of participants in which the option that was previously the compromise is now displayed as either a high or low extreme. For example, imagine in our case instead of the gamble above you were *initially* shown the set

	Option		
	A	**B**	**C**
Attribute 1	50ps	25ps	0ps
Attribute 2	5kz	8kz	10kz

If you had indeed initially been shown this set, which option would be the rational choice? The answer is not the 50/5 option that was the compromise in the last set, but, if one is risk averse, the 25/8 option that is the compromise alternative in this new set. The reason is this: *Because the attribute scales have no absolute metric in a payoff (or utility) spac*e, and since no learning is possible, the normative analysis is exactly the same for both the old and new choice sets. An assumption that individuals are risk averse over uncertain preferences leads to a predicted bias toward choosing the middle option on a Pareto frontier in both cases.

Hence, what the compromise effect shows is *not* that people do not make choices in a way that is consistent with utility theory, but rather that *one assumption* that accompanies typical applied analyses may not always hold. That is, the assumption that attribute levels have a constant absolute meaning (and mapping to utility) across a range of choice sets. If one relaxes that assumption and builds a model that formally recognizes uncertainty in utility exchange rates, compromise data can be easily reconciled with standard theory. The effect *seems* anomalous only because readers have access to holistic knowledge about the range of attributes that participants in the experiments did not.

Are the models suggested by Kivetz et al. wrong? To the contrary, the above analysis ironically leads us to the same conclusion we reached about how one can algebraically *describe* compromise effects. That is, their contextual concavity model (expression 8) can be directly motivated as a model of risk-averse preferences for consumers who are unsure about the scaling of attributes and subjectively normalize them over the range displayed in the experiment. The critical difference is that by explaining the result in terms of its origins in risky decision making, we can endogenously predict the model's failure. Expression (10) implies that preferences for compromise options should vanish as uncertainty in preferences diminishes and/or individuals are exposed to choice sets with broader ranges sequentially.

Conclusion

It would be wrong to conclude that the above illustration implies that rational models of choice enjoy any higher status as tools of explanation than any other type of model (behavioral or statistical), or that behavioral researchers should reject all possible normative benchmarks (however far-fetched) before deriving their own explanation. In this case, a rational model of risky choice was appropriate for entirely pragmatic reasons; it offered a convenient representation that satisfied the theoretical modeling requirements of simplicity and endogenous recognition of uncertainty. Were we to continue this modeling effort, ideally the flow of dialogue would again reverse, with behavioral research being asked to refine the risky choice model to better reflect the realities of how consumers actually deal with preference uncertainty and to incorporate lay beliefs about the scaling and benefits of attributes (e.g., Machina 1982).

As we mentioned at the outset of the paper, our goal was to foster more dialogue among proponents of what have become increasingly disparate approaches to understanding and modeling decisions and choices. In recent years, behavioral research has made significant advances in providing better understanding of how consumers learn preferences and make choices, but without

a clear connection to empirical choice modeling, either in terms of how efforts can be improved by this knowledge or what the consequences are of ignoring it. In this paper we tried to illustrate in a limited way how to begin to build such bridges. We hope it represents only the first words in what should prove to be a long dialogue.

Acknowledgments

The authors are listed alphabetically. Jordan J. Louviere is professor of marketing and acting executive director at the Centre for the Study of Choice (CenSoC), Faculty of Business, University of Technology, Sydney, Australia. Robert J. Meyer is the Gayfryd Steinberg professor of marketing and co-director of the Center for Risk and Decision Processes at the Wharton School of the University of Pennsylvania, Philadelphia. The authors thank Eric Bradlow and Gal Zauberman for their helpful comments on an earlier draft of the manuscript.

Notes

1. Expression (2) can be seen as a special case of the general class of case-based decision models described by Gilboa and Schmeidler (1995), where the similarity function is defined over the maximum-similarity referent stored in memory.
2. That is, recover the preferences that would be revealed under full information.
3. This follows from the definition of concavity, which requires $[v(.5k) + v(.5k)] > [v(0k) + v(1k)]$.

Appendix 1.1. Individual Logit Estimation Results

Tables 1A.1 through 1A.4 follow.

Table 1A.1

Individual-Level MNL Model Estimates for Flights

ID	MSR	4 hrs	5 hrs	6 hrs	7 hrs	$350	$450	$550	$650	Qant	Vblue	Jstar	Aust	FFlyer	Stops	Drinks	ASC1	ASC2
1	0.662	–0.060	–0.150	0.200	–0.452	–0.150	0.220	0.000	–0.070	1.870	0.760	–0.900	–1.730	0.090	–0.050	0.050	1.630	1.350
2	0.864	0.080	0.050	0.040	–0.994	1.770	0.860	–0.410	–2.220	0.150	–0.230	0.000	0.080	0.220	0.240	–0.230	1.370	1.520
3	0.791	0.410	–0.030	–0.130	–1.171	1.900	0.690	–0.340	–2.250	–0.310	0.220	–0.010	0.100	0.010	–0.030	0.130	1.480	1.350
4	0.534	0.000	0.000	0.000	–0.534	1.880	0.860	–0.710	–2.030	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.600	1.300
5	0.534	0.000	0.000	0.000	–0.534	1.880	0.860	–0.710	–2.030	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.600	1.300
6	0.806	0.290	–0.030	–0.290	–1.066	1.730	0.710	–0.760	–1.680	–0.110	0.100	–0.050	0.060	–0.100	0.120	–0.010	1.750	1.510
7	1.014	0.760	0.020	–0.070	–1.794	1.820	0.830	–0.680	–1.970	–0.470	0.290	0.380	–0.200	–0.030	–0.030	0.090	1.550	1.300
8	1.235	0.800	0.960	–0.180	–2.995	1.170	1.020	–0.530	–1.660	–0.120	0.120	–0.010	0.010	0.130	–0.020	–0.280	1.200	1.590
9	1.692	–0.070	0.330	0.380	–1.952	0.640	0.470	–0.420	–0.690	0.250	1.400	–2.720	1.070	0.070	0.040	0.040	1.110	1.350
10	0.534	0.000	0.000	0.000	–0.534	1.880	0.860	–0.710	–2.030	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.600	1.300
11	1.310	0.640	0.010	–0.510	–1.960	1.610	1.310	–0.570	–2.350	0.390	–0.080	0.010	–0.320	–0.080	0.080	–0.340	1.410	1.510
12	0.924	–0.200	–0.100	–0.110	–0.624	1.770	0.730	–0.850	–1.650	–0.180	–0.200	0.000	0.380	0.180	–0.150	–0.090	1.830	1.400
13	0.534	0.000	0.000	0.000	–0.534	1.880	0.860	–0.710	–2.030	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.600	1.300
14	1.520	0.700	0.330	–0.800	–2.550	0.590	0.580	–0.030	–1.140	–0.160	0.020	–0.040	0.180	–0.110	0.570	0.010	2.320	1.720
15	0.534	0.000	0.000	0.000	–0.534	1.880	0.860	–0.710	–2.030	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.600	1.300
16	1.022	1.000	0.070	–0.160	–2.092	1.670	1.020	–0.840	–1.850	–0.460	0.280	0.180	0.000	0.030	0.090	–0.040	1.360	1.470
17	1.057	0.500	0.230	–0.430	–1.787	1.690	0.670	–0.860	–1.500	–0.010	0.270	–0.010	–0.250	0.000	0.230	–0.200	1.690	1.790
18	1.038	0.630	–0.190	–0.210	–1.478	1.090	0.280	–0.070	–1.300	0.200	0.190	–0.370	–0.020	–0.080	0.290	–0.680	1.830	1.660
19	0.911	1.410	0.920	–0.490	–3.241	1.100	–0.050	–0.320	–0.730	–0.230	0.000	0.150	0.080	0.000	–0.240	–0.080	1.570	1.480
20	2.197	0.200	0.050	0.050	–2.447	0.160	0.600	–0.220	–0.540	0.510	0.990	–1.940	0.440	0.010	0.610	–0.530	1.430	2.000
21	0.614	–0.080	0.300	–0.120	–0.834	1.500	0.810	–0.500	–1.810	0.270	0.060	–0.090	–0.240	–0.140	–0.300	–0.350	1.520	1.690

Abbreviations: MSR = mean squared residual; Qant = Qantas; Vblue = Virgin Blue; Jstar = JetStar; Aust = Australian Airlines; Fflyer = Frequent Flyer program; Stops = number of stops; Drinks = free wine/beer; ASC = alt-specific constant; hrs = flying time; $ = fare.

Table 1A.2

Individual-Level MNL Model Estimates for Pizzas

ID	MSR	Phut	Dom	Boys	Phav	$12	$14	$16	$18	ntop1	ntop2	ntop3	ntop4	bread	drink	desrt	ASC1	ASC2
1	2.287	0.700	0.200	–0.480	–3.187	0.820	0.590	0.670	–2.080	–1.310	0.260	0.770	0.280	–1.020	–0.130	–0.220	1.280	1.060
2	2.519	0.280	–0.380	0.010	–2.419	0.650	0.620	0.080	–1.350	–1.510	0.560	0.070	0.880	–0.530	–0.570	–0.290	1.790	1.740
3	1.279	–0.740	0.180	0.010	–0.719	0.420	0.440	0.570	–1.430	–1.630	–0.170	0.840	0.960	–0.230	–0.260	0.000	1.710	1.570
4	2.127	–0.610	0.010	1.120	–1.527	–0.510	0.610	–0.040	–0.060	–1.980	0.070	1.350	0.560	0.050	–0.260	–0.430	1.560	1.530
5	1.750	–0.300	0.300	0.000	–1.750	0.400	–0.020	–0.240	–0.140	–1.680	–0.480	0.830	1.330	–0.220	–0.660	0.180	1.840	1.160
6	2.932	0.100	–0.080	–0.710	–2.952	0.050	–0.080	0.200	–0.170	–3.020	–0.100	1.510	1.610	0.110	–0.260	0.030	1.360	1.040
7	3.186	–0.770	0.730	0.290	–3.146	–0.180	0.470	0.310	–0.600	–2.430	0.190	1.120	1.120	–0.430	–0.800	–0.520	1.590	1.240
8	2.310	1.210	2.310	–3.060	–5.830	–0.230	0.000	0.000	0.230	0.090	0.210	–0.230	–0.070	–0.150	0.150	–0.100	0.880	0.920
9	1.310	–0.070	–0.210	0.170	–1.030	1.100	0.920	–0.350	–1.670	–1.040	1.030	0.210	–0.200	–0.380	0.000	–0.330	1.830	1.250
10	2.358	–0.340	–0.060	–0.160	–1.958	–0.080	0.280	–0.030	–0.170	–2.700	–0.580	1.030	2.250	–0.320	–0.200	0.050	1.030	0.920
11	1.974	–0.290	–1.000	–0.080	–0.684	–0.320	–0.080	0.240	0.160	–0.190	0.130	0.400	–0.340	–0.210	–0.340	–0.140	2.110	2.000
12	2.635	0.530	–0.280	–0.320	–2.885	0.190	0.330	0.250	–0.770	–2.410	0.630	0.850	0.930	–0.860	–0.280	–0.290	1.560	0.980
13	2.193	–0.320	0.270	–0.180	–2.143	0.480	1.040	0.120	–1.640	–1.310	0.070	0.490	0.750	–0.450	–0.470	–0.090	1.490	2.210
14	2.366	–0.470	1.950	0.540	–3.846	–0.370	–0.580	0.180	0.770	–0.030	–0.490	–0.440	0.960	0.220	0.340	–0.140	1.220	1.580
15	1.901	1.180	1.910	–0.820	–4.991	0.170	–0.050	–0.120	0.000	–0.420	–0.140	–0.120	0.680	0.060	–0.370	–0.200	1.350	1.540
16	0.919	–0.160	0.220	0.170	–0.979	1.340	0.990	–0.310	–2.020	–0.910	0.660	0.420	–0.170	–0.100	–0.090	–0.060	1.600	1.570
17	2.402	–2.270	0.860	1.140	–0.992	–0.010	0.030	0.300	–0.320	–1.570	–0.080	0.950	0.700	0.000	–0.010	–0.410	1.270	1.690
18	1.043	–0.420	2.040	0.750	–2.663	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.370	1.300
19	1.084	–0.030	–0.480	0.300	–0.574	1.090	0.060	0.130	–1.280	–0.010	–0.030	0.160	–0.120	–0.260	–0.070	–0.760	1.900	1.520
20	2.850	–0.250	–0.190	–0.150	–2.410	0.460	0.660	0.130	–1.250	–1.670	–1.370	1.260	1.780	0.020	–0.340	–0.150	1.520	1.050
21	1.232	–0.280	0.110	0.160	–1.062	1.010	1.060	0.290	–2.360	–1.150	–0.020	0.890	0.280	–0.270	–0.300	0.060	1.490	1.510

Abbreviations: MSR = mean squared residual; Phut = Pizza Hut; Dom = Domino's; Boys = Eagle Boys; Phav = Pizza Haven; ntop = number of toppings; bread = free bread; drink = free drinks; desrt = free dessert; ASC = alt-specific constant; $ = price.

Table 1A.3

Summary Statistics for Individual MNL Models: Flights

Effect	N	Mean	StdError	T-Stat	Skewness	StdError	Kurtosis	StdError
Residua1^2	21	0.968	0.095	10.202	1.325	0.501	1.918	0.972
Time, 4hrs	21	0.334	0.095	3.525	0.862	0.501	0.076	0.972
Time, 5hrs	21	0.132	0.066	1.991	1.946	0.501	3.508	0.972
Time, 6hrs	21	–0.135	0.057	–2.362	–0.693	0.501	1.211	0.972
Time, 7hrs	21	–1.434	0.193	–7.429	–0.585	0.501	–0.811	0.972
Fare $350	21	1.403	0.134	10.439	–1.370	0.501	0.910	0.972
Fare $450	21	0.717	0.066	10.940	–0.799	0.501	1.403	0.972
Fare $550	21	–0.521	0.060	–8.751	0.648	0.501	–0.710	0.972
Fare $650	21	–1.598	0.137	–11.644	1.079	0.501	0.314	0.972
Qantas	21	0.076	0.105	0.719	2.687	0.501	9.752	0.972
Virgin Blue	21	0.200	0.086	2.312	1.946	0.501	3.783	0.972
JetStar	21	–0.258	0.161	–1.600	–2.646	0.501	6.726	0.972
AusAirlines	21	–0.017	0.107	–0.161	–1.710	0.501	8.365	0.972
FreqFlyer	21	0.010	0.020	0.480	0.655	0.501	0.498	0.972
# stops	21	0.069	0.049	1.417	1.032	0.501	1.469	0.972
Drinks	21	–0.120	0.046	–2.585	–1.324	0.501	1.244	0.972
ASC1	21	1.574	0.054	28.952	0.993	0.501	3.343	0.972
ASC2	21	1.485	0.043	34.753	1.079	0.501	0.752	0.972

Table 1A.4

Summary Statistics for Individual MNL Models: Pizzas

Effect	N	Mean	StdError	T-Stat	Skewness	StdError	Kurtosis	StdError
Residua1[2]	21	2.031	0.145	13.964	–0.225	0.501	–0.969	0.972
Pizza Hut	21	–0.158	0.160	–0.985	–0.554	0.501	2.891	0.972
Domino's	21	0.400	0.199	2.009	0.992	0.501	0.075	0.972
Eagle Boys	21	–0.062	0.186	–0.333	–2.092	0.501	7.472	0.972
Pizza Haven	21	–2.274	0.310	–7.346	–0.963	0.501	0.717	0.972
$12	21	0.309	0.117	2.638	0.386	0.501	–0.867	0.972
$14	21	0.347	0.098	3.557	0.019	0.501	–0.662	0.972
$16	21	0.113	0.056	2.026	0.168	0.501	0.274	0.972
$18	21	–0.769	0.195	–3.951	–0.224	0.501	–1.136	0.972
# toppings = 1	21	–1.280	0.204	–6.264	–0.065	0.501	–0.859	0.972
# toppings = 2	21	0.017	0.110	0.151	–0.604	0.501	2.058	0.972
# toppings = 3	21	0.589	0.120	4.905	–0.196	0.501	–0.934	0.972
# toppings = 4	21	0.675	0.154	4.382	0.449	0.501	–0.300	0.972
Free bread	21	–0.237	0.067	–3.515	–1.007	0.501	1.128	0.972
Free drinks	21	–0.234	0.058	–4.035	–0.066	0.501	0.426	0.972
Free dessert	21	–0.181	0.049	–3.722	–0.830	0.501	0.840	0.972
ASC1	21	1.512	0.065	23.411	–0.119	0.501	0.090	0.972
ASC2	21	1.399	0.077	18.172	0.498	0.501	–0.140	0.972

References

Abe, Makoto. 1995. "A Nonparametric Density Estimation Method for Brand Choice Using Scanner Data." *Marketing Science* 14 (3): 300–325.

Allenby, G.M., and P.E. Rossi. 1999. "Marketing Models of Consumer Heterogeneity." *Journal of Econometrics* 89 (1–2): 57–78.

Bettman, J.R., E.J. Johnson, and J.W. Payne. 1991. "Consumer Decision Making." In *Handbook of Consumer Behavior*, ed. T.S. Robertson and H.H. Kassarjian. Englewood Cliffs, NJ: Prentice-Hall.

Bradlow, Eric, Y. Hu, and T.-H. Ho. 2004. "A Learning-Based Model for Imputing Missing Levels in Partial Conjoint Profiles." *Journal of Marketing Research* 41 (November): 369–381.

Camerer, Colin, and Teck Ho. 1999. "Experience-Weighted Attraction Learning in Normal Form Games." *Econometrica* 67 (1999): 837–874.

Dawes, Robin, and Bernard Corrigan. 1974. "Linear Models in Decision Making." *Psychological Bulletin* 81: 105–106.

Einhorn, H.J. 1970. "The Use of Nonlinear, Concompensatory Models in Decision Making." *Organizational Behavior and Human Performance* 73: 221–230.

Einhorn, H.J., D.N. Kleinmuntz, and B. Kleinmuntz, 1979. "Linear Regression *and* Process-Tracing Models of Judgment." *Psychological Review* 86: 465–485.

Elrod, T., R.D. Johnson, and J. White. 2005. "A New Integrated Model of Noncompensatory and Compensatory Decision Strategies." *Organizational Behavior and Human Decision Processes* 95 (1): 1–19.

Erdem, Tülin, Susumu Imai, and Michael P. Keane. 2003. "A Model of Consumer Brand and Quantity Choice Dynamics under Price Uncertainty." *Quantitative Marketing and Economics* 1 (1): 5–64.

Erdem, Tülin, and Michael P. Keane. 1996. "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets." *Marketing Science* 15 (1): 1–20.

Fudenberg, Drew, and David K. Levine. 1998. *The Theory of Learning in Games.* Cambridge, MA: MIT Press.

Gilboa, Itzhak, and David Schmeidler. 1995. "Case-Based Decision Theory." *Quarterly Journal of Economics* (November): 605–630.

Gilbride, T.J., and G.M. Allenby. 2004. "A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening." *Marketing Science* 23 (Summer): 391–406.

Gonul, Fusun, and Kannan Srinivasan. 1996. "Estimating the Impact of Consumer Expectations on Coupons on Purchase Behavior: A Dynamic Structural Model." *Marketing Science* 15 (3): 282–280.

Gregan-Paxton, Jennifer, and Deborah Roedder John. 1997. "Consumer Learning by Analogy: A Model of Internal Knowledge Transfer." *Journal of Consumer Research* 24 (3): 266–287.

Hensher, David A., and William H. Greene. 2003. "The Mixed Logit Model: The State of Practice." *Transportation* 30: 133–176.

Houser, Daniel, Michael Keane, and Kevin McCabe. 2004. "Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence Using a Bayesian Type Classification Algorithm." *Econometrica* 72 (3) (May): 781–822.

Hutchinson, J. Wesley, and Robert J. Meyer. 2005. "Extreme Bias: Heuristics and Biases in Intuitive Judgments about Maximum-Order Statistics." Working paper, Department of Marketing, the Wharton School, University of Pennsylvania.

Johnson, E.J., R.J. Meyer, and S. Ghose. 1988. "When Choice Models Fail: Compensatory Models in Negatively Correlated Environments." *Journal of Marketing Research* 26 (3): 255–270.

Kamakura, W.A., B.D. Kim, and J. Lee. 1996. "Modeling Preference and Structural Heterogeneity in Consumer Choice." *Marketing Science* 15 (2): 152–172.

Kamakura, W.A., and M. Wedel. 2004. "An Empirical Bayes Procedure for Improving Individual Level Estimates and Predictions from Finite Mixtures of Multinomial Logit Models." *Journal of Business and Economic Statistics* 22 (1): 121–126.

Keeney, R.L., and H. Raiffa. 1976. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs.* New York: Wiley.

Kivetz, R., O. Netzer, and S. Srinivasan, 2004, "Alternative Models for Capturing the Compromise Effect," *Journal of Marketing Research* 41(3): 237–257.

Louviere, J.J. 1988. *Analyzing Decision Making: Metric Conjoint Analysis.* New York: Sage.

Louviere, J.J., and T.C. Eagle. 2006. "Confound It! That Pesky Little Scale Constant Messes Up Our Convenient Assumptions!" Proceedings, Annual Sawtooth Software Conference, Delray Beach Florida, March (http://www.sawtoothsoftware.com/techpap.shtml).

Louviere. J., D. Hensher, and J. Swait. 2000. *Stated Choice Methods: Analysis and Applications.* London: Cambridge University Press.

Machina, Mark. 1982. "Expected Utility Analysis Without the Independence Axiom." *Econometrica* 50 (March): 277–323.

Manchanda, Puneet, Peter E. Rossi, and Pradeep Chintagunta. 2004. "Response Modeling with Nonrandom Marketing Mix Variables." *Journal of Marketing Research* 41 (4): 467.

McFadden, Daniel. 1981. "Econometric Models of Probabilistic Choice." In *Structural Analysis of Duscrete Data with Econometric Applications*, ed. C.F. Manski and D. McFadden, 198–272. Cambridge, MA: MIT Press.

Mellers, Barbara E. 1981. "Configurality in Multiple-Cue Learning." *American Journal of Psychology* 93: 429–443.

Meyer, Robert J., 1987. "The Learning of Multiattribute Judgment Policies," *Journal of Consumer Research* 14 (September): 155–173.

Meyer, Robert J., and Howard Kunreuther. 2005. "Earthquakes in the Lab: The Optimality of Decisions to Invest in Long-Term Mitigation." Working paper, Center for Risk and Decision Processes, University of Pennsylvania.

Norman, Donald A. 1988. *The Psychology of Everyday Things*. New York: Basic Books.

Payne, J.W., J.R. Bettman, and E.J. Johnson. 1993. *The Adaptive Decision Maker.* New York: Cambridge University Press.

Rossi, Peter E., and Greg M. Allenby. 2003. "Bayesian Statistics and Marketing." *Marketing Science* 22 (3): 304–328.

Rossi, P.E., G.M. Allenby, and R. McCulloch. 2005. *Bayesian Statistics and Marketing.* Hoboken, NJ: John Wiley and Sons.

Sheng, Shibin, Andrew M. Parker, and Kent Nakamoto. 2005. "Understanding the Mechanism and Determinants of Compromise Effects." *Psychology and Marketing,* 22 (7): 591–610.

Song, Inseong, and Pradeep Chintagunta. 2003. "A Micromodel of New Product Adoption with Heterogeneous and Forward-Looking Consumers: Application to the Digital Camera Category." *Quantitative Marketing and Economics* 1 (4): 371.

Swait, Joffre, and Wictor Adamowicz. 2001. "The Effect of Task Complexity on Consumer Choice: A Latent-Class Model of Decision Strategy Switching." *Journal of Consumer Research* 28 (June): 135–148.

Swait, J., and J.J. Louviere. 1993. "The Role of the Scale Parameter in the Estimation and Use of Multinomial Logit Models." *Journal of Marketing Research* 30: 305–314.

Ter Hofstede, F., Y. Kim, and M. Wedel. 2002. "Bayesian Prediction in Hybrid Conjoint Analysis." *Journal of Marketing Research* 34 (2): 253–261.

Tversky, A. 1972. "Elimination by Aspects: A Theory of Choice." *Psychological Review* 79 (4): 281–299.

Tversky, A., and I. Simonson. 1993. "Context-Dependent Preferences." *Management Science* 39 (10): 1179–1189.

Wedel, Michel, and Wagner A. Kamakura. 1999. *Market Segmentation: Conceptual and Methodological Foundations.* 2d ed. Boston: Kluwer Academic Publishers.

CHAPTER 2

HOW MUCH TO USE?

An Action-Goal Approach to Understanding Factors Influencing Consumption Quantity

VALERIE S. FOLKES AND SHASHI MATTA

Abstract

Factors influencing how much an individual consumes on a single-usage occasion are identified drawing on research in consumer behavior as well as allied disciplines. The overarching framework for understanding how various factors influence the amount consumed is based on Gollwitzer's (1996) "action-goal" model. Initially, such factors as a product's price and social norms influence the perceived desirability and feasibility of consumption-related goals. In the next phase, such factors as self-control strategies and product instructions influence the selection of a means of implementing the goal. However, during execution, individuals can be distracted from their planned means of implementing the goal. Finally, consumers' motivation to use feedback, and the type of feedback about consumption, influences subsequent goal setting. The integrated framework we propose provides a means to understand how, and at what stage, various factors affect usage quantity. Such an understanding can aid marketers in formulating products, designing packaging, and creating messages, and can help public policy makers identify effective strategies to manage the well-being of consumers and of the environment.

A consistent and striking feature of the consumer behavior literature is its focus on purchase rather than on consumption (e.g., Helgeson et al. 1984; Holbrook and Hirschman 1982; Kassarjian 1978; Wells 1993). Although consumption is often conditional on purchase, a science of consumer behavior that does not follow through to an examination of consumption itself must necessarily be considered incomplete. Further, when inquiry is rooted in a particular context—as when the knowledge base for consumer behavior is derived mainly from the purchase context—the variables important to that particular context can dominate and limit theoretical development in the larger domain of interest.

The study of how and why people consume products generally centers on factors that include the person (who consumes), the product (what is consumed), the temporal dimension (when people consume), the locus of consumption (where people consume), or the quantity consumed (how much). Although the factors are interrelated and each is important to investigate, we focus on the quantity consumed in this review. Understanding how and why consumers make usage quantity

decisions has important practical as well as theoretical implications that warrant scientific inquiry. Usage quantity decisions are ubiquitous, with costs for individuals and society. From the moment individuals awaken in the morning, they make decisions about how much of a product to use. Consumers make mundane decisions about how much shampoo to apply, how much water and coffee to combine to make the perfect beverage, and how much coolant to put in the car's radiator, as well as more portentous decisions about how much of a pharmaceutical product to consume, how much fat-laden food to eat, and how much alcohol to drink before driving.

Marketing managers are interested in the amount consumed because one way to increase product sales and hence profit is by increasing the amount used by individual consumers. Marketers may also want to influence the quantity consumed because of its link with customer satisfaction. If consumers use amounts that are perceived to yield benefits (which could involve either an increase or a decrease in amount), satisfaction will increase, and with it corresponding gains in customer loyalty.

Usage quantity issues are a public policy concern because they can have major implications for consumers' safety and well-being (e.g., a high amount of sugar consumption can lead to diabetes for some people, heavy smoking contributes to birth defects and cardiovascular disease). Environmentalists' assignment of usage quantity as the first of the "three Rs" (reduce, reuse, and recycle) establishes it as a social issue. Persons intent on the conservation of natural resources maintain that even small changes in the amount consumed by each individual can have large effects in the aggregate.

Influencing the total amount consumed requires analysis of factors affecting the two components of usage—the frequency of use and the amount consumed on each occasion. The purpose of this chapter is to identify what is known about how consumers decide how much of a product to use on single consumption occasions (as opposed to the total quantity used across occasions or the frequency of usage), as well as the gaps in that knowledge. Most previous research examining usage quantity does so within the context of a particular social issue (e.g., obesity, energy conservation, alcoholism). A broader perspective is taken here to identify commonalities in usage quantity issues across heretofore conceptually segregated bodies of research because insight into problems identified in one field may facilitate understanding in a seemingly disparate field.

First, a framework of usage decisions is presented that provides conceptual coherence in understanding how people make such decisions. The analysis of usage quantity decisions is grounded in recent advances that relate individuals' goal setting to their goal-striving actions. Then research in consumer behavior as well as research in allied disciplines that address usage issues is reviewed and evaluated in light of the proposed model.

Overview of the Action-Goal Model Applied to Consumers' Usage

The overarching framework for understanding how various factors influence the amount used is based on the analysis of action goals by Gollwitzer (1996). The "action-goal" model provides a means of understanding when and how various factors influence a particular act. Although not formulated with consumption in mind, the model is applicable to understanding the amount consumed.

Phases of Action

Gollwitzer's (1996) model identifies the sequence in which cognitions about a goal relate to the performance of tasks to achieve the goal. Consumption goals are achieved through four quali-

Table 2.1

Usage Phases and Examples for Energy Consumption, Food Consumption, and Pharmaceutical Consumption

Phase	Task	Energy	Food	Pharmaceutical
Predecisional	Identify goals	Compare effort in conserving energy to convenience	Compare health benefits to taste benefits	Compare quality of life if regimen is followed versus status quo
Preactional	When, where, and how of starting	Install appliance	Create a menu	Read prescription instructions
Executional	Actual product usage	Operate appliance	Ingest food	Apply or ingest medication
Postactional	Evaluate outcome of goal striving	Compare electricity bills	Assess health	Analyze side effects

tatively distinct tasks, two of which involve goal setting and two of which involve goal striving. The sequential ordering consists of four phases: predecisional, preactional, executional, and postactional (Table 2.1).

The model assumes that consumers have a variety of wishes or desires. Consumers' task in the predecisional phase is goal setting - to identify which goals they will strive to achieve. They must analyze the desirability and feasibility of various wishes or options. For example, a consumer might evaluate the desirability of reducing home energy use. Factors influencing the desirability of a goal can be internal or external to the consumer. For example, goal desirability is likely to be context dependent, so social pressures can influence the incentives to pursue a consumption goal. Similarly, the perceived feasibility of achieving a consumption goal can be influenced by internal and external factors. For example, consumers may differ in their sense of self-efficacy, which would influence their beliefs about the feasibility of reducing energy use.

Having set a goal, the consumer proceeds to a goal-striving phase. The task of this preactional stage is "the when, where, and how of getting started" (Gollwitzer 1996, p. 289). Guides to selecting a course of action for consumption can be internal to the consumer or external. For example, product instructions direct consumers to use various amounts, but consumers might instead rely on their previous experience to develop a plan.

In the other goal-striving phase that follows, the consumer executes the behavior, actually consuming the product. Various factors can facilitate the consumer's pursuit of a course of action, but sometimes distractions can disrupt the execution of behavior. For example, a consumer might intend to limit alcohol intake, but situational influences could distract the individual from that plan. Again, internal and external factors play a role.

Following the executional phase, goal-setting issues are again salient in the postactional phase. The task is to evaluate the outcome of goal striving, which influences the subsequent selection of goals. Consumers can invest varying amounts of effort in evaluating the outcome of usage. For example, a consumer might analyze the efficacy and the anticipated side effects of using more than the prescribed dose of medicine, or might give little thought to the consequences. Additionally, usage situations differ in the type and amount of feedback easily accessible to the consumer.

In sum, the amount used by consumers during consumption can be influenced at any one of the four phases of goal-directed consumption. Each phase involves a different task for successful goal completion (Gollwitzer 1996). Hence, a particular factor may influence usage at different points for different reasons, sometimes increasing the amount consumed and other times decreasing the

amount consumed. For example, social influence can enhance the desirability of a goal (phase 1) and later be a distraction from adhering to a course of action (phase 3). The influence of a factor at one stage is contingent on the successful completion of a previous stage, indicating a way in which one factor may influence another. More detail about the factors influencing usage quantity is provided in the discussion of the individual phases.

Phases at Which Consumption Quantity Is Considered

Although Gollwitzer's (1996) action-goal model provides a useful means of identifying the tasks involved in decisions about consumption, it does not imply that decisions about how much to use are always considered in setting a goal during the first phase. Rather than considering how much to use when setting a goal, the consumer might decide how much to use when planning a course of action to take when using the product. For example, a consumer could decide to treat a cold with a certain medicine and then read the instructions to decide how much of the medicine to take. Further, although the consumer may decide before action initiation how much to eat (e.g., to eat half a hamburger and drink a small cup of coffee), consumption quantity decisions often are unnecessary for the planning of action initiation in phase 2. A person can make usage quantity decisions while consuming.

The fact that consumption can often be initiated without deciding how much to use has implications for the implementation of consumption quantity decisions. A consumer who has not made a quantity decision before initiating consumption may invest little effort in developing predictions about effects of various quantities, and have vague goals about usage quantities with little commitment to those goals (cf. Bagozzi and Edwards 1998). Quantity decisions made "on the fly" in phase 3 may be more subject to distracting environmental influences than when made prior to initiating consumption. For example, a dieter need not relate the quantity of food consumed to the weight loss goal to order a restaurant meal. The dieter's plan to order a low-calorie plate would serve the weight loss goal, but lack of a plan about the quantity to eat may facilitate substitution of a more salient sensory goal about the amount to eat.

If vague plans have been established about usage quantities, consumers will have difficulty identifying useful feedback from monitoring amounts. For example, the decision to eat two pieces of candy is easily monitored by counting candy wrappers, but criteria for satisfying one's goal to eat "some" bread are less clear. In sum, consumers' motivation to invest cognitive resources may be lower for usage quantity decisions than for consumption decisions pertaining to what, where, and when to consume.

Summary

Having established a general framework for understanding usage quantity decisions, we can now consider specific factors influencing the quantity of consumption at each phase of the model and examine empirical evidence for their influence. Note that research related to one phase can also be relevant to another. Our chapter also identifies issues that warrant further exploration. Because relatively few consumer behavior studies address usage, much of the following discussion synthesizes knowledge from diverse sources and fields on what may seem to be disparate topics.

The Predecisional Phase (Phase 1)

The goal-setting task of the predecisional stage is to evaluate the desirability and feasibility of acting on a goal (Gollwitzer 1996). That evaluation can transform mere wishes and desires to a binding

intention. The intention "specifies a desired end-state, which may be the execution of a concrete behavior or the attainment of a desired outcome" (Gollwitzer 1996, p. 292). For example, goals related to the consumption of alcohol can be identified on an abstract level (to enjoy oneself) or on a concrete level (to have only one drink). Hence, a wide variety of goals can relate to consumption activities. These goals may conflict, as when the "restrained drinker" is caught in a cyclical pattern of consumption (restraint or binging) depending on the fluctuating desirability of conflicting goals (e.g., cravings for alcohol versus self-esteem and self-presentational goals; Bensley 1991).

Previous research on consumption quantity has explored the effect of a specific product attribute (price) on the desirability of consumption. Other research has investigated the impact of social influence on goal setting. Research on product ownership effects suggests additional influences on the desirability and feasibility of consuming certain quantities.

Product Price

Price is an extrinsic property of products that can influence the desirability of consuming. Consumers may infer a product's desirability from its price and so want to consume more of it. However, economic costs generally increase with the quantity consumed so that the product's perceived price acts to inhibit consumption (Wansink 1996). Women indicate they would use more of even such inexpensive products as Crisco, M&M's, and Creamette Spaghetti when they believe the price is low. Similarly, they would use more of Mr. Clean and Crisco when those products are offered on sale than when they are regularly priced. Hence, even small amounts of money appear to have a direct influence on amounts consumed.

Of course, price incentives can also increase the quantity purchased, which can, in turn, influence consumption goals. Ailawadi and Neslin (1998), using scanner panel data, modeled the effect of price promotions on the rate of consumption in two product categories (yogurt and ketchup). Promotions were found to increase household inventory levels, which in turn increased consumption. This may be at least partly because reducing inventory becomes a desirable goal.

Though price promotions or price reductions may intuitively suggest greater purchase quantity, this relationship is not as simple as it seems. Wertenbroch (1998) argues that consumers purchase different quantities of a product when there is a reduction in unit price (e.g., a quantity discount), depending on whether the product is a vice product (e.g., cigarettes, fatty foods) or a virtue product (e.g., low-fat items). His research contends and finds support for the hypothesis that, all else being equal, consumers are less price sensitive for vices than for virtues, and hence would forego savings from price reductions through quantity discounts for vices. Results suggest that consumers impose constraints over purchase (and subsequent consumption) of vice products by effectively paying price premiums in order to engage in self-control. Though this research does not directly test for whether the reduced purchase quantity of vice products reduces consumption quantity of those products, it shows that the nature of the product can affect the quantity purchased.

Price influences depend on the context of usage or the way costs are framed. A product's price is a simple means of calculating the cost of consumption, but people can be influenced to exert effort to calculate other costs. Energy conservation increases when consumers calculate the total expenditures for using an appliance over the life cycle of the appliance (Hutton and Wilkie 1980). Framed in that way, the immediate savings from buying a cheaper but energy-inefficient appliance can seem trivial.

Context also influences price perceptions in that the salience of purchase costs (e.g., price, shopping effort) probably declines as the temporal interval between purchase and usage increases. Over time consumers would be likely to forget prices (e.g., Dickson and Sawyer 1990). The

exterior protective packaging and other cues to the purchase occasion (e.g., price tags) would be discarded. Other costs (e.g., storing the product) and benefits (e.g., immediate sensory gratification) would become more salient. For example, discounts on large sizes of products make them attractive purchases in warehouse club stores, but large units may be inconvenient for home use. In short, consumers' perceptions of costs when using may vary considerably depending on the salient cues in the usage context. Because of the effort often required to retrieve price information and then integrate it with other, more accessible information, price is likely to have less impact on consumption than it has on purchase.

Further, price may have a more complex effect on the amount consumed than the research suggests when it is perceived as an extrinsic incentive to consume. A consumer's perceived source of motivation to consume influences the desirability of consumption and so is likely to influence the amount consumed. Whether consumption is perceived as intrinsically or extrinsically motivated alters one's pleasure in consumption. Persons who experience autonomy over their behavior like and persist in the behavior more than those whose behavior is controlled by external forces (for a review, see Deci and Ryan 1987). Factors undermining a sense of autonomy include extrinsic incentives to consume, threats of aversive consequences from engaging in consumption, deadlines for completion of activities, surveillance by others, and lack of perceived choice over outcomes.

Social Information

Whereas research examining extrinsic consumption incentives suggests that others' influence attempts can lower the enjoyment of consumption, research also indicates that social influence can enhance the desirability of consumption. Information about others' consumption influences the perceived desirability and feasibility of a goal. Worchel, Lee and Adewole (1975) found that products made scarce because of others' desire for them are evaluated more favorably than items offered in constant supply. They manipulated a supply of cookies so that it was large (10 cookies), small (2 cookies), or diminishing (10 cookies of which 8 were subsequently withdrawn). After tasting the cookies, students indicated that they wanted to eat more of them and believed they were more desirable when they were in limited supply, particularly when the supply diminished from 10 to 2.

Consumption by others suggests the amount that is feasible to consume, so that conforming with consumption norms becomes a goal. Homeowners who were given information about energy expenditures by residents in their area reduced their consumption by more than 20 percent in comparison with those who did not receive that information (Van Houwelingen and Van Raaij 1989). Others' usage seems to have provided a reference point against which the homeowners could evaluate their own usage as excessive, although it also may have made competition with others on the usage dimension a salient and desirable goal.

Effects of Owning Products on the Desirability of Consuming

Although not directly examining amounts consumed, research on ownership indicates that it may influence goal setting. The desirability of consuming may differ depending on whether the product is owned or unowned. Although involvement may decrease after purchase (Richins and Bloch 1986), ownership can induce a greater liking for even inexpensive products (Beggan 1992), which in turn seems likely to increase the pleasure of consuming and the amount consumed. Consumption can sometimes provide a means of incorporating a product into the self, thus affording a sense of possession (cf. Belk 1988). A greater quantity of consumption may heighten such a perception.

On the other hand, one consequence of usage is necessarily a decrease in one's supply. People may be reluctant to relinquish valued products (cf. Kahneman 1992). Whether a greater liking for a product will increase consumption of it or increase reluctance to relinquish possession of it and thus lead to restraint in the quantity consumed is unclear. Perhaps the effect depends on the source of meaning that gives an object value (e.g., utilitarian versus self-expressive, see Richins 1994).

Other Factors Likely to Be Important in the Predecisional Phase

Consumers' personality dispositions are likely to influence their evaluations of consumption goals. Some traits, such as compulsive buying ("chronic and repetitive purchasing that becomes a primary response to negative events or feelings"; O'Guinn and Faber 1989, p. 155) and materialism (valuing acquisition and the means to acquire possessions; Richins and Dawson 1992), suggest a predisposition to view consumption and consuming as desirable goals. Individuals with a higher need for uniqueness value scarce commodities as a means of differentiating the self (Snyder 1992), perhaps also leading to a greater amount consumed than normative. Other traits, such as impulse control, learned industriousness (Eisenberger 1992) and self-efficacy (Bandura 1977), may relate to the perceived feasibility of goals.

The effect of the consumer behavior context (purchase versus consumption) on the perceived desirability of goals has not been explored, but is likely to influence the quantity consumed. Consider the potential differences in the comparison of alternatives. The salient alternatives in a purchase context are other brands of the same product, some of which may not be in one's consideration set (e.g., Pan and Lehmann 1993; Ratneshwar, Shocker, and Steward 1987). In contrast, in a usage context, alternatives are generally limited to the stock on hand, and hence comparisons are likely to involve dissimilar products. Such a comparison could alter the salience of a product's attributes and the consumers' evaluation and subsequent usage of the product (cf. Bettman and Sujan 1987; Johnson 1989; Simonson and Winer 1992).

More research is needed into cultural differences in consumption quantities and cultural trends. Values in Western culture have emphasized status acquired via conspicuous consumption of quantities of goods. Yet, some have noted a shift from status linked to quantity of consumption, both symbolic and material, to quality (Shipman 2004). Conspicuous consumption in the future might be less likely to be manifested by the ability to waste than by a spare or simplified lifestyle that expresses taste.

The Preactional Phase (Phase 2)

Phase 2 is distinguished from phase 1 by its orientation toward implementing goals rather than setting goals. Planning for the initiation and implementation of a task can be simple when the actions are well practiced (Gollwitzer 1996). Some simple action sequences, such as rituals, will have rich meaning embedded in them closely associated with important goals (cf. Rook 1985). In contrast to those who are secure in their routines, the individual who is uncertain of how or where to act invests more cognitive effort in planning for the initiation and implementation of action (Gollwitzer 1996). Sometimes that cognitive effort is devoted to extensive search and complex evaluation before engaging in action. For example, a consumer may attempt to recall nonexperiential knowledge from books and manuals as well as knowledge gained from experience. Hence, the preactional stage includes simple and complex cognitive activity.

Previous research has investigated several aspects of goal implementation in this stage. Some has addressed the link between goal setting and goal implementation by demonstrating that con-

sumption quantity can be a means of implementing nonconsumption goals. Other research has explored strategies of self-control over usage and product-specific beliefs. Several studies have investigated the use of external guides (e.g., instructions) to consumption.

Consuming Amounts to Achieve Nonconsumption Goals

Decisions about the amount to consume can be a means of implementing nonconsumption goals, as is evident from the literature examining social influence and food consumption. Consumption amounts are perceived as a means of conveying achievement of social ideals. College students believe they will appear more feminine by eating less and appear more masculine by eating more (Chaiken and Pliner 1987). Women in particular eat less when they are with an attractive man than they do when they are with another woman (Mori, Chaiken, and Pliner 1987; Pliner and Chaiken 1990). Even if others are not present at the time, women eat less when hungry if another person will be able to identify how much they have eaten (Polivy et al. 1986). Men, in contrast, eat more in groups than they do when alone (Edelman et al. 1986). Additionally, men have more concerns about social censure and about rejection arising from consuming too little alcohol than women do (Teahan 1987).

Women's concern about the impressions created by the amount of food consumed is well founded. College students perceived women who ate a small breakfast and a small lunch as more feminine than women who ate larger meals (Chaiken and Pliner 1987). Impressions of men did not differ with meal size.

Stereotypes about persons who consume too much are not limited to eating, nor is gender always a moderator of such stereotypes. People hold stereotypes about persons who differ in energy-conserving behavior (Sadalla and Krull 1995). Although energy conservation tends to be relatively private, self-presentational concerns may influence the quantity consumed. Some of the stigma attached to using more rather than less may be due to blaming the individual for not exercising restraint. For example, attributing obesity and alcoholism to lack of control leads to condemnation of people with those conditions (for a review, see Weiner 1986). In sum, the research suggests that the amount consumed can be influenced by seemingly irrelevant goals. Perhaps people use whatever means are at their disposal to implement important goals, such as goals of self-presentation.

Preactional Means of Exerting Self-Control over Consumption

Implementation of consumption goals may require strategies that help the consumer regulate usage. Often such goals involve forgoing the immediate rewards from consumption because of long-term costs, as when people refrain from eating too much food or drinking too much alcohol to avoid future health problems. People may also need self-control to consume amounts of products that have long-term benefits but few immediate ones (e.g., eating bland diet food) and to time the rewards of usage to receive maximal effect (e.g., savoring products by distributing one's consumption across occasions). Research examining those situations has identified some strategies consumers use to regulate their usage.

Abstinence

One means of reducing the amount used is abstinence. "The technology of self-control often implies outright prohibitions because allowing a little bit eventually leads to excesses" (Thaler

1985, p. 208). For example, Alcoholics Anonymous maintains that abstinence is the only means of controlling alcohol intake. Although typical weight control strategies involve eating less food, certain high-calorie foods generally are prohibited. Research suggests that once a forbidden food has been consumed, the dieter no longer maintains restraint (Knight and Boland 1989).

Identifying conditions that trigger a course of action can help an individual refrain from consumption. One condition is the individual's affective state. Negative moods appear to diminish one's ability to resist temptation (Fry 1975). For example, bulimics who are aware that a bad mood is a common precondition of binge eating tend to refrain from a course of action they know will lead to negative feelings (Fairburn 1993). Passing up the initial opportunity to engage in an activity has the additional advantage of leading to continued forgoing of the activity (Tykocinski, Pittman, and Tuttle 1995).

Precommitment Strategies

Various precommitment strategies also aid in self-control over consumption. As consumption often cannot occur unless purchase preceded it, the type and the amount of product purchased influence the amount consumed. For example, purchase of an energy-efficient appliance facilitates conservation, but purchase of an energy-inefficient appliance results in overconsumption of energy (Anderson and Claxton 1982). Consumers can exercise control over consumption (1) by not purchasing a desired product, (2) by purchasing less of a desired product, or (3) by purchasing product variants that make usage costly or less desirable.

The numerous product variants in the marketplace that make usage costly or less desirable suggest that many consumers opt for them as a means of controlling quantities (e.g., filter cigarettes, reduced-fat foods). A potential disadvantage of this strategy is that it might not reduce the quantity used and could even increase it. Use of a substitute or ersatz product could make risks less salient, removing inhibitions against usage.

Savoring

Self-control is exercised not only when consumers perceive negative consequences in consuming, but also when they temporally regulate the amount consumed to derive maximal benefits. People prefer to separate positive outcomes, enjoying them more by experiencing them on separate days rather than at one time (Linville and Fischer 1991; Thaler and Johnson 1990). For example, students prefer to receive news of excellent grades on each of two papers on separate days rather than on the same day.

Although not tested in the context of product consumption, the principle of separation of gains seems to hold true for food products and perhaps other products as well. Consuming the same food repeatedly decreases its desirability. In one study, ratings of a chocolate candy's tastiness declined at a faster rate after several pieces had been eaten than the rate of decline that occurred after the candy color was changed between pieces (Rolls, Rowe, and Rolls 1982). That finding suggests an individual might perceive greater total sensory rewards from eating (and savoring) one piece of a favored item each day than from eating several pieces at one time.

Goal Execution Beliefs in the Preactional Stage

Research suggests that consumers develop product-specific beliefs that guide implementation of goals in the peractional stage. People hold "naïve" theories about the causal relationship between

symptoms and medications that can influence the amount and frequency of medicine usage (Pennebaker and Watson 1988). Some people with high blood pressure believe they can estimate their blood pressure from their symptoms and emotions and plan their medical regimen accordingly. However, their blood pressure estimates are no more accurate than the estimates of those who lack confidence in their ability to estimate their blood pressure. People also appear to hold theories about satiety, relating energy to food attributes (Booth 1994). Drinkers develop more positive and fewer negative expectancies about effects of ingesting quantities of alcohol than nondrinkers (Grube, Ames, and Delaney 1994). Many college-aged women relate the quantity of alcohol consumed to declines in social inhibitions (Teahan 1987).

In short, product experience affects consumers' naïve theories about cause-and-effect relationships about quantities. Few studies have explored the ways in which consumers test and develop their hypotheses. Consumers may use feedback about product performance to arrive at notions about the effects of various quantities through a trial-and-error process, or may use a more top-down approach, applying schemas and heuristics. For example, smokers may find through a trial-and-error process that adverse consequences do not immediately follow from each cigarette, and might thus conclude that they can dismiss warnings of dangers from smoking (Breznitz 1984).

Besides developing product-specific beliefs, consumers may rely on fairly general heuristics about products and amounts to make plans about how much to consume. Such heuristics may be common because they decrease the effort required to plan for usage. Forming plans for usage can be complex if the individual must take into account multiple causal factors as well as multiple effects (e.g., food intake, rate of consumption, and amount all influence the effects of alcoholic beverage consumption).

Consumption-quantity heuristics have not been explored in the literature. Yet people seem to apply usage heuristics that relate usage quantities to potency, persistence, and onset of effects. The notion that "more is better" suggests a linear relationship between amount and effect leading to the conclusion that overuse will yield better effects. However, some conditions may elicit the reverse belief—that less is better. One study found that many physicians treating pulmonary tuberculosis prescribe insufficient amounts of drugs, thus imposing considerable financial costs for salvaging therapy (medical regimens that compensate for previous mistreatment) and causing detriment to the patients' health (Mahmoudi and Iseman 1993). Presumably, physicians underprescribed in the belief that better outcomes would follow. The notion that "more is longer" suggests a temporal relationship between amount and effect, in that using more will lead to persistence of effects. Similarly, "more is quicker" suggests that using more will lead to more immediate effects. Research is needed to investigate heuristics and beliefs about amounts and the conditions that elicit them.

Also related to consumers' preactional beliefs or theories is the notion of tradeoff between consumption choices. Consumption of one product may affect the choice and consumption of another product in the same episode, as when a consumer plans consumption not just by focusing on single products but also within the larger consumption context or "consumption episode" (e.g., the choice of an entrée and dessert during a meal) (Dhar and Simonson 1999). Consumers sometimes engage in a tradeoff between a goal (e.g., pleasure) and a resource (e.g., money), or a tradeoff between two goals (e.g., pleasure and good health). Consumption strategies seem to differ in these two cases. If engaging in a tradeoff between goals and resources, consumers choose a highlighting strategy (e.g., a tasty, expensive appetizer and a tasty, expensive entrée on one occasion, and a less tasty, less expensive items on another occasion). When engaging in a tradeoff between two goals, consumers choose a balancing strategy (e.g., a healthy entrée and a tasty dessert, or a tasty entrée and a healthy dessert). That research, though it does not address consumption quantity specifically, is relevant to understanding how consumption quantity of one product

can affect the consumption quantity of another product during the same consumption episode. Consumption can depend on the nature of tradeoffs that consumers engage in (e.g., consuming a large meal on a special occasion and consuming a smaller meal on another more regular occasion versus consuming a smaller quantity of the entrée in order to consume a larger quantity of dessert during the same meal).

The Impact of External Guides to Usage

Companies and government agencies often provide information through instruction labels and warnings about when, where, and how to use products that can help people implement goals. Yet research suggests that such information has little impact on behavior (for a review on warnings, see Stewart, Folkes and Martin 2001). Consumers typically perceive external guides to executing usage decisions as having little value.

Consumers commonly maintain that they will not read instructional material on products (Wright 1981; Wright, Creighton, and Threlfall 1982). Survey respondents report that they do not read product owner's manuals because they believe they already know how to use the product, they lack the time to read them, and they believe they can learn about the product more quickly by using it (Celuch, Lust, and Showers 1992). Those explanations may reflect some broader beliefs. First, consumers appear to believe their generalized knowledge about product use is sufficient when they consider the effort involved in acquiring narrow, brand-specific knowledge. Second, they appear to believe misuse entails few risks and prescribed use entails insubstantial benefits. Such a belief fosters simplified decision making as well as being a possible consequence of simplified decision making (e.g., lack of attention to product performance and feedback about usage). Finally, consumers appear to prefer trial as a means of information acquisition, as suggested by previous research comparing how product trial and other types of marketing communications influence consumers (e.g., Hoch and Deighton 1989; Kamins and Assael 1987; Smith 1994).

Whereas consumers appear to lack motivation to search for instructional information, they will use information when little search is necessary. Energy conservation studies indicate that labels attached to devices as reminders to reduce usage have varying degrees of effectiveness depending on their attention-attracting properties. In a field study, small signs and stickers on classroom light switches did not reduce energy usage, but large signs with bold lettering at classroom exits did (Winett and Kagel 1984). Those guides may have been successful because the students did not make the link between their own energy conservation goal and turning off the classroom lights until the sign drew attention to it.

Even if consumers read instructional information, they may lack the ability or motivation to comply with it. Consider the well-known Food Guide Pyramid. Even after being instructed on the use of the Food Guide Pyramid in selecting serving sizes, students still have difficulty understanding the concept, though they performed better than those without instruction (Knaust and Foster 2000).

Misinterpretation of external information can lead to important errors. Wansink and Chandon (2006) suggest that "low fat" nutritional labels increase food intake by increasing perceptions of appropriate portion size and by decreasing guilt associated with consumption. They contend that consumers' perceptions of serving size are unreliable and vary by as much as 20 percent. In three experiments, a low-fat label increased consumption of a snack (M&M's or granola) compared to when there was no such label, and low-fat labels made consumers feel less guilty about how much they consumed. Moreover, overweight consumers underestimated the calorie content of a

low-fat snack significantly more than normal weight consumers. Even normal weight consumers have difficulty understanding information about food serving sizes.

Even if consumers have the ability to understand compliance instructions, they might not be motivated to comply. Consumers seem to question the trustworthiness of instructions provided by firms. Consumer skepticism could arise because firms often have obvious incentives to encourage overuse (as when increased usage increases profits). However, firms' usage appeals seem to have little impact even when they advocate reduced consumption. A study examining techniques to encourage energy conservation found that brochures explaining how to reduce energy use were more effective in gaining compliance when a government body rather than the local power company urged conservation (Craig and McCann 1978). Consumer overuse in response to firms' usage recommendations may occur for many types of products, as when consumers use more than the recommended amount of products, inferring that fear of legal repercussions leads pharmaceuticals manufacturers to be overly cautious. In short, consumers' inferences about motives behind firms' recommendations may decrease compliance with instructions (cf. Kirmani and Wright 1989; Wright 1986).

Gaining compliance with instructions has been a problem even in situations when motivation to comply should be fairly strong and the instruction giver's credibility is high. Most medical therapies address problems that elicit high-involvement information processing from patients. Further, the information source is more credible than most sources of product instructions because the consumer's welfare is the main concern of medical personnel and prescriptions are tailored to the patient's needs. Although some studies have shown compliance with medical regimens to be 95 percent, others have found it to be less than 40 percent, even low enough that the patient's health is at risk (Dunbar-Jacob, Dwyer, and Dunning 1991).

Some of the noncompliance may arise from lack of comprehension. Research suggests that people may have difficulty understanding common instructions (Mustard and Harris 1989). When given sixteen different actual prescription labels, about half of a group of college students interpreted them incorrectly (Mustard and Harris 1989). The results seem to suggest that instructions should always be as simple as possible.

However, other research has shown that the simplest instructions do not always elicit the greatest compliance. The fit between the consumer's goals when using the product and the form of instructions influences compliance (Martin and Folkes 2001). Novice product users comply with more complex instructions about amounts to use when they are motivated to maximize product outcomes than they do when they want to minimize the effort in using products. Novice product users who are motivated to minimize effort comply with simple instructions about amounts to use more than those who want to maximize product performance. Hence, the fit of the particular type of instruction with the consumer's goal in using the product influences compliance. A good fit also enhances the consumer's confidence in using the product and satisfaction with the product.

Another reason for noncompliance may be patients' tendency to substitute their own judgment based on experience with a regimen for experts' advice. For example, alcohol-warning labels are effective in reducing the alcohol intake of pregnant women who have had no previous live births but not for women who have had one or more live births (Hankin et al. 1996).

In sum, consumers are more likely to follow product-related consumption guides when the instructions capture attention and are comprehensible, when the consumers lack or perceive themselves to lack product experience, and when the consumers are motivated to comply. Generally, consumers appear to lack motivation to utilize instructional information, but might use it more when it reduces guilt associated with overuse. Less cognitively demanding instructions should facilitate compliance with labels and instructions (e.g., pictorial instructions or ones that integrate the prescribed quantity, as when cold pills are bubble packed by twos).

Other Factors Likely to Be Important in the Preactional Phase

More research is needed to examine the motivational component of quantities consumed, especially in regard to affectively laden quantity-of-usage phenomena (e.g., savoring, hoarding, or gorging). Additionally, the way people relate settings to products needs additional investigation. People seem to have well-developed schemas about the situations in which products are used (e.g., soup is eaten for lunch but not breakfast; Wansink and Ray 1996). They may also hold schemas about products and settings that regulate consumption quantities. Consumption-related beliefs that influence the quantity of consumption include perceptions of products as complementary, interchangeable, or incompatible. For example, wine and beer are often perceived as interchangeable yet incompatible for simultaneous or sequential consumption, and the norms about amounts to consume differ. Understanding beliefs about relationships among products should shed light on decisions about the when, where and how of consuming.

The Execution Phase (Phase 3)

The task in the third phase is executing goal-directed behaviors (Gollwitzer 1996). The situation may offer opportunities that facilitate goal achievement or present distractions from pursuing the course of action to completion. Usage can involve highly overlearned and repetitive behaviors enacted under familiar conditions with little incentive to monitor and revise the actions. Such situations increase the likelihood that the chosen action will be carried through to completion. Factors that facilitate overuse in the execution stage include having a large supply of the product on hand, increasing the salience of the product, modifying the product or package to encourage usage and creating consumption atmospherics that encourage usage.

Despite the desire to adhere to a course of action, the consumer may encounter unanticipated obstacles or unsought feedback requiring revisions of executional decisions and repeated search and reevaluation. Further, underuse seems more likely to occur in the execution stage than overuse. Increased use of a product commonly requires additional effort or intensity of performance and persistence in a course of action. Factors that increase the effort needed to use more of the product or interfere with persistence in usage are likely to reduce the quantity consumed.

Available Supply

The supply on hand is an obvious factor restricting the amount used. The supply places an upper bound on the amount of product that can be consumed because an individual cannot use more than the amount available at any one time. Hence, purchase of a large amount permits more usage than the purchase of a small amount. However, the supply also influences usage in less obvious ways. A small supply of a product might cue a variety of cognitions about goal execution, such as thoughts about replenishment costs, predictions for future amounts to consume, and the formation of intentions for replenishment (e.g., a low reading on a gas gauge makes gas prices salient, leading to predictions about when refueling will be necessary).

Perhaps because of these thoughts, consumers apportion a smaller amount of a product for usage when a small supply is available than when a large supply is available (Folkes, Martin, and Gupta 1993). Given a partially full but large container, they put smaller amounts of detergent into a washing machine, apply smaller amounts of shampoo to their hair, and pour less cleanser into a bowl than they do when the same sized container holds twice as much of the product. The total supply rather than the fill level of the container accounts for the difference. Consumers allocate

the same amount of the product when the supply is held constant but the size of the container differs (e.g., pour the same amount when an 8 oz. container is full and when a 16 oz. container is half full). Moreover, the supply effect can not be explained by differential perceptions of product price, potency, and quality associated with fill level or container size. People did not associate a small bottle or a small supply with a higher-quality, more potent, or more expensive product. Further, it is not due to difficulties in regulating the amount during the act of pouring the product from the container. However, visual assessment of the supply seems essential because when the container was opaque, the supply effect disappeared. Perhaps the visual cue increases the salience of the supply as a cue in apportioning an amount to use.

One's supply can be viewed as a resource that consumers become increasingly reluctant to give up as it diminishes. An exception to that rule seems to occur when the consumer has a supply remaining that is slightly greater than the amount that would be apportioned for a single usage occasion. Then, a larger amount is used in an apparent attempt to finish the product and avoid retaining an incomplete portion (Folkes, Martin, and Gupta 1993). For example, a person might scoop a generous portion of ice cream from a nearly empty container rather than leave a minute amount remaining. The reason could be that immediate, salient concerns (storage costs, perceptions of product efficacy) outweigh future costs (inventory replenishment costs).

Another limitation to the effect of supply on usage is a ceiling effect for large supplies. In one of their experiments, Folkes, Martin and Gupta (1993) gave students about to launder their clothes a large container of detergent to use. Students used less detergent when the bottle was one-third full compared to when it was two-thirds full. However, they used about the same amount when the bottle was two-thirds full as when it was full. It may be that a ceiling effect emerged because students feared overusing detergent would harm their clothing. Hence, the tendency to use more with a greater supply occurs only within a limited range of amounts.

Product Salience

Just as the supply of a product can be a salient cue for usage, so also can the salience of the product itself. Product salience is another perceptual facilitator that can increase consumption quantity. Research has shown that making food products visible can stimulate unplanned consumption, even when consumers are satiated. Cornell, Rodin, and Weingarten (1989) allowed their respondents to eat a meal (sandwiches, fruit, potato chips, and brownies) to satiety, before exposing them to pizza. The diners were asked to rate their desire to eat pizza, taste the pizza, and then allowed to eat as much pizza as they wanted. Results showed that respondents' intentions to eat pizza were higher when it was in front of them (compared to pre- and post-lunch ratings). Despite being satiated, they consumed the pizza.

Similarly, visual salience of supply facilitates consumption of stockpiled foods. Chandon and Wansink (2002, study 2) gave consumers either large or moderate quantities (stockpiling versus nonstockpiling) of different kinds of high-convenience or low-convenience foods, and increased the salience of certain foods by providing pictures of them. They then monitored household consumption over a twelve-day period. Overall, stockpiled foods were consumed more than nonstockpiled foods, and high-convenience foods (ready-to-eat foods) were consumed twice as fast when they were stockpiled. Chandon and Wansink (2002) argued that recently stockpiled foods can be visually salient and that could be one reason why they are consumed more.

Product salience seems to particularly influence consumers who find it difficult to restrain consumption. Nederkoorn and Jansen (2002) exposed high- and low-restraint eaters to food, while physiological measurements were taken. Subsequently, respondents participated in a taste

test during which food intake was measured. Unrestrained eaters showed an increase in heart rate, gastric activity, and saliva during food exposure, suggesting that they prepared for food intake, while no such activity was evident in restrained eaters. Gastric activity also significantly correlated with consumption quantity.

Finally, product sampling can facilitate consumption because it primes or induces a motivation to consume more of a high-incentive product (e.g., tasty food). Wadhwa, Shiv, and Nowlis (2006) show that sampling a tasty food or drink enhances consumers' subsequent consumption of other tasty food or drinks. In three experiments they find that consumers who sampled a tasty beverage (study 1) or snack (study 2) consumed more food or drink in a subsequent task than consumers who did not sample (study 1) or those who did not sample as tasty a product (study 2). Further, they find that an individual difference variable, Behavioral Activation System (BAS) sensitivity, moderates this effect. Consumers high on the BAS scale were more likely to subsequently consume a greater amount of cola and more likely to rate other hedonic products favorably if they sampled a sports beverage initially, than if they did not sample the beverage. Those authors argue that BAS is a motivational system that underlies behavioral response tendencies such that individuals with a high BAS sensitivity have a higher motivation to seek out appetitive behaviors when primed with a high-incentive cue.

Package Shape and Product Amount Perceptions

Package shape can influence consumption quantity through perceptions of package size. This effect appears to arise from perceptual biases. Raghubir and Krishna (1999) found that consumers use the height of the container or its elongation to simplify volume judgments of the product inside. A container's height predicted respondents' volume judgments better than or about as well as models that included width or depth. When containers were tall or elongated, respondents in their studies perceived those containers as having more of a product than those that were shorter or squat in shape. If consumers perceive elongated containers to contain more of the product, this increase in perceived available supply should influence the consumption quantity of that product (recall Folkes, Martin, and Gupta 1993). Indeed, Wansink and van Ittersum (2003) conducted research with teenagers at weight-loss camps and with nondieting adults and found that the elongation bias caused teenagers to pour and drink 88 percent more juice or soda into short, wide glasses than into tall, narrow glasses that held the same volume. These teenagers underestimated the amount they poured by as much as 50 percent. Similarly, when experienced bartenders were asked to pour 1.5 ounces of an alcoholic beverage into short and wide glasses, the bartenders on average poured more than when they poured into tall and narrow glasses.

The elongation effects may sometimes arise because the elongated products examined in these studies seem to have attention-attracting properties. Package shapes that attract more attention are also perceived to contain a greater volume of a product than same-sized packages that attracted less attention (Folkes and Matta 2004). The disparity in attention seems to lead to "mental contamination" of the volume judgment. If the containers are similar in size, the one that attracts attention more is judged larger. Folkes and Matta argue that the bias is due to the covariation of the attention directed to a package with its size. Through experience with a variety of stimuli, people probably have learned to determine which of two shapes is larger when the magnitude of the difference is large. A perceptual sensation that covaries with that size judgment is differential attention—large shapes generally attract attention more than small shapes. This covariance of attention and size causes the overall attention-attracting properties of a package to bias or contaminate volume judgments such that packages that attract more attention are also perceived to contain a greater

volume of a product than same-sized packages that attract less attention. Because package shapes that attract more attention lead to perceptions of greater available supply, those perceptions should in turn increase consumption quantity when compared to the consumption quantity of a product contained in a less attention-attracting package (cf. Folkes, Martin, and Gupta 1993).

Product Design "Affordances" Influencing the Execution Stage

Consumers can also purchase products with design properties that facilitate adherence to one's course of action. The design aspects of products have been termed "affordances"—"the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used" (Norman 1988, p. 9). Influences on quantity consumption include the features or affordances that initiate use and terminate use.

Norman (1988) analyzed the effect of product design on the implementation of a goal to conserve natural resources. Most home water faucets require one action to start usage and the reverse of that action to stop usage. The action sequence initiating use is generally physically and temporally distinct from the action terminating use. Hence, such faucets do not facilitate water conservation. In contrast, water faucets found in many public facilities allow water to flow only when pressure is applied to hold the faucet in position. The initiating action must be sustained with some effort, whereas ceasing the effort terminates usage. The basic difference in design has an obvious consequence for water conservation. Products requiring exertion of effort to initiate use and cessation of effort to terminate use will lead to lower consumption than those for which continued usage requires no effort.

Product design also facilitates the calibration of amounts so that they are consistent with one's intentions. "Garbology" studies examining the kind and amount of refuse discarded by individual households find considerably more waste when people buy prepared foods than when they buy fresh foods that can be more finely calibrated into portions at purchase (Rathje 1992).

More research is needed on product design issues, a subject rarely addressed in the marketing literature (Bloch 1995).

Consumption Atmospherics

The intensity of sensory factors such as lighting, odor, and noise in the consumption environment seem to influence consumption quantity in an inverted U-shaped fashion, though sometimes the absence of controls makes this conclusion conjectural. Dimmed or soft lighting, when compared to harsh or bright illumination in restaurants, increases eating duration and also increases comfort and disinhibition (Lyman 1989; Ragneskog et al. 1996), hence contributing to increased consumption. An inverted U-shaped relationship between lighting and the amount consumed seems probable because very dim lighting seems likely to decrease the salience of food.

Rolls and Rolls (1997) find that food odors can stimulate consumption of a food, but may also decrease consumption of that food if people experience sensory-specific satiety when exposed to the odor of that food for a period of time (experiment 2). An implication of this finding is that within a single meal, consumers' intake of food can be limited if presented with the same odor for a period of at least several minutes before the eating starts or during the initial part of the meal. Therefore, slow eating, which allows olfactory and gustatory sensory-specific satiety to build up, may tend to reduce consumers' meal size. Conversely, odor variety in a meal can enhance consumption of food.

Music is another atmospheric variable that seems to influence the amount of food consumed.

Liked music appears to encourage longer meal duration and greater consumption of food in a restaurant compared to no music or disliked music (Caldwell and Hibbert 2002; North and Hargreaves 1996, 1998). Similarly, Milliman (1986) suggests that background music tempo influences the rate and quantity of consumption in a restaurant.

Perhaps these sensory effects are partly due to alterations in mood that lead consumers to revise their beliefs about products and time allocations to eating. Consumers in a positive mood have improved expectations about products that are not evaluated negatively (e.g., candy would create a positive mood, which would in turn improve expectations about the consequences of consuming it; Kahn and Isen 1993). More favorable product expectations might increase the amount consumed so that once consumers experience mood elevation from consuming the product, they would use more than they anticipated before initiating consumption.

Social Facilitators

Consumers can facilitate adherence to their selected course of action by creating usage environments that reward such activities. Research shows that social support from one's family is an important factor in influencing whether a patient complies with a medical regimen (Aaronson 1989; Dunbar-Jacob, Dwyer, and Dunning 1991). Such support makes following through on the action easier and so facilitates goal striving in the execution phase.

Other social factors facilitating consumption include the presence of others and social modeling. Eating with familiar people can lead to an extended meal and greater consumption of food (cf. Bell and Pliner 2003; de Castro 1994; de Castro and Brewer 1992). In a study that had respondents maintain a seven-day detailed diary of everything they ate or drank and those consumption episodes, de Castro (1994) found that meals eaten with other people were larger and longer in duration compared to meals eaten alone. The relationship with one's companion also influenced consumption. Meals eaten with spouse and family were larger and eaten faster than with others. Meals eaten with friends were larger and of longer duration than with less intimate others. Those results suggest that the presence of others at a meal increases consumption by extending the duration of the meal, and that familiar others have a stronger effect, probably by providing a relaxed atmosphere and disinhibition. In another study, de Castro and Brewer (1992) found that meals eaten with one other person were, on average, 33 percent larger than those eaten alone, and that meals eaten with seven other people present were as much as 96 percent larger than those eaten alone. Similarly, Bell and Pliner (2003), in an observational study of consumers in a variety of settings (a worksite cafeteria, a fast-food restaurant, and a moderately priced restaurant), found a significant positive relationship between meal duration and the number of people eating at each table.

On the other hand, consuming food with unfamiliar others can reduce consumption quantity in situations where self-monitoring is high (e.g., first dates, job interviews, etc.) (cf. Pliner and Chaiken 1990; Mori, Chaiken, and Pliner 1987). Women targets who consumed relatively small meals were rated as more feminine (Chaiken and Pliner 1987) and more socially appealing (Basow and Kobrynowicz 1993) than their counterparts who consumed larger amounts of food. Further, research has also shown that women tend to consume less food in the presence of others, particularly in the presence of attractive males (Mori, Chaiken, and Pliner 1987), suggesting the effect of self-presentational concerns on consumption quantity. Indeed, social modeling has been shown to enhance consumption of water (Engell et al. 1996, experiment 2) such that respondents who were in the presence of a confederate drinking a large amount of water consumed a larger amount of water when compared to those who were alone or in the presence of a confederate who consumed a smaller amount of water.

Social disapproval, even when it comes from a stranger, can stop an individual from a consumption habit. In one study, when signs forbidding smoking in a cafeteria were accompanied by verbal requests to stop smoking, the mean minutes spent smoking fell dramatically from the level observed when only signs were posted (Jason and Liotta 1982).

Effort Exertion

More effort required to use the product generally decreases consumption, though whether the user perceives effort exertion positively or negatively can depend on the user's intrinsic interest in the activity. People can define tasks in ways that increase their intrinsic interest in the activity and therefore their consumption (Sansone et al. 1992). An activity that is conducive to embellishment in a fantasy context can generate increased interest (Parker and Lepper 1992). For example, when the use of a car care product is associated with a race-car driver, the user might elaborate on a racing fantasy, consequently enjoying using the product more and valuing the effort invested in consuming the product.

Effort, when conceptualized as the difficulty or inconvenience in accessing food products, has been shown to affect the amount of food that is consumed by consumers. Overall, studies have found that lower effort increases consumption. Engell at al. (1996) showed that respondents drank more water when the water was available on the dining table versus when it was 20 or 40 feet away (experiment 1). Painter, Wansink, and Hieggelke (2002) examined the consumption of candy in a study of secretaries who were given access to candy either on their desks or at a distance of 2 meters away from their desks. They found that, on average, when the candies were placed on their desks, the secretaries ate about 5 more candies than when they had to exert more effort, that is, stand up and walk 2 meters to eat them.

Hearn et al. (1998) examined data from two school nutrition projects and found that availability and accessibility of fruits and vegetables at school and home increased school children's consumption of fruits and vegetables. Wing and Jeffery (2001) find that consumers who were given low-calorie food provisions and menus were more likely to consume those low-calorie foods compared to consumers in other study conditions (control group and other intervention techniques). They conclude that reducing the effort in having access to low-calorie foods increases diet compliance and weight loss.

Distractions from Usage Execution

Effort plays a role in another factor influencing the execution of a usage decision. Distraction from using a product can influence the amount consumed. The effect of a distractor can differ with the task. When effort must be exerted to terminate consumption, a repetitive consumption cycle leads to overconsumption. Research suggests that concentrating on nonconsumption tasks that require information-processing effort increases repetitive consumption rather than distracting from it. When engaging in a task that required their constant attention, people ate more snacks as the task increased in difficulty (Cantor, Smith, and Bryan 1982).

Distracting Product Features

Product features that interrupt the behavioral chain of consumption should decrease consumption amounts more than those that do not interrupt usage. The beginning of the action sequence of

consumption provides cues that remind the user of the alternative of termination. The division of a product into small units seems likely to prompt more frequent assessment of the quantity to use. Kirschenbaum and Tomarken (1982) examined serving size effects by offering unlimited amounts of ice cream but varying the size of the bowl in which it was served. Their women subjects ate less when given small bowls than when given large ones, perhaps because the decision to stop was cued more frequently by the small bowl.

Some studies suggest that variety or lack of product uniformity can act as a distractor, increasing the amount of food consumed at a single meal. In other words, product variety in an assortment can affect consumption quantity. In some of the early psychological research on the effect of variety on consumption quantity, Rolls, van Duijvenvoorde, and Rolls (1984) gave participants a four-course meal that was either varied or homogenous (one each of sausages, bread and butter, chocolate dessert, and bananas, or four courses of one of these foods), and found that respondents given the varied meal consumed 44 percent more food than those who received the same food for each course. Similarly, when offered a succession of sandwiches for lunch, students ate 33 percent more when the sandwiches had four different fillings than when they all had the same filling (Rolls et al. 1981). In another study in that series, respondents consumed significantly more grams of yogurt when provided with three flavors over successive courses as compared with participants provided with the same flavor of yogurt in all courses (Rolls et al. 1981). Moreover, mere changes in the shape of food increase the amount ingested. When students were offered in succession a variety of pasta shapes (bow ties, hoops, and spaghetti), they ate more than they did when they were offered only their favorite shape (Rolls, Rowe, and Rolls 1982).

These effects of variety on consumption could be moderated by cognitive processing difficulty. Kahn and Wansink (2004) identified three moderators—organization, symmetry, and size of assortment—that affect consumption. In one study, they gave respondents an assortment of M&M's candies in either seven or ten different colors, with taste being identical among all colors. They also manipulated symmetry of the assortment by varying the frequency of one color (10 percent vs. 30 percent dark brown). Increasing the number of colors of M&M's candies increased respondents' consumption of the candies only when the assortment was asymmetric. This occurs because asymmetric distributions are easier to process compared to symmetric distributions. In an asymmetric assortment the dominant items are processed and appreciated first, before the remaining items. On the other hand, when the assortment is symmetrically distributed, there is no easy heuristic for processing the variety. Respondents who received the high variety (10 colors) rated the candies more fun to eat when the assortment was asymmetric (30 percent brown) than when symmetric (10 percent brown). Hence, increasing assortment variety may increase consumption only when the assortment is not too complex.

The tendency to consume more of heterogeneous products may be limited to ones with sensory properties, such as food (sensory-specific satiety; see Rolls, Rowe, and Rolls 1982). However, a desire for stimulation might contribute to such an effect for nonfood products. Product heterogeneity (e.g., multicolored toothpaste) may evoke the consumer's curiosity, reducing self-imposed constraints on consumption and encouraging impulsivity (cf. Loewenstein 1994).

Distraction in the Environment

Distraction can occur not only as a result of product attributes, but also as a result of participating in multiple tasks at the same time. Research has shown that such distraction can affect an individual's ability to monitor the quantity consumed (Bellisle and Dalix 2001). This research found that healthy women who listened to a detective story when eating (i.e., distracted respondents) consumed more

food than did healthy women who ate alone (baseline group), those who ate alone and listened to instructions on the sensory characteristics of food, or those who ate in groups of four. Respondents' cognitive restraint as indicated by their monitoring of satiety was found to correlate with the difference in the food consumed between the baseline and distraction conditions.

Cognitive distraction affects the attention individuals pay to oral sensory signals when consuming food. In a study conducted by Poothullil (2002), lean female respondents ate cereal when they felt hungry under baseline and three other conditions: eating until the pleasantness of the flavor of the cereal subsided, eating until the stomach felt full, and eating while watching television. Results showed that cereal consumption was significantly higher than baseline when respondents ate until they felt full and while watching television. Recognition of oronasal sensory cues while eating was offset by cognitive distraction in monitoring satiety, leading to an increase in consumption quantity.

Similarly, Stroebele and de Castro (2004) found that respondents in a two-week diary study reported being less hungry when they ate meals or snacks while watching television than when they were not watching television. Although the meals with the television on were smaller, the frequency of meals increased. They concluded that normal internal cues regulating food intake may not be as effective while watching TV (p. 113). This is consistent with prior research that suggests that people consume more food and snacks when watching television and may do so even when not actually hungry (Tucker and Bagwell 1991; Tucker and Friedman 1989).

The Postactional Phase (Phase 4)

The task in the final, postactional phase is evaluation of goal achievement (Gollwitzer 1996). Feedback may reveal that the outcome of the course of action has fallen short of one's desires. Consequently, standards of performance may be lowered or alternative goals pursued.

The evaluative processes in the final phase often involve elaborative inference, explanation, and prediction, which can form a distinct knowledge base from the practical knowledge associated with implementation ("how to" use the product; Ohlsson 1996). Acquiring, interpreting, and using feedback can be complex. Consumers may need to monitor amounts as they are being consumed to determine whether consumption meets their goals and to know how much has been consumed in the past if the past is to be used as a standard in the present. When a person is drawing complex inferences about why amounts are effective, knowledge of the effects of using certain amounts can influence consumption (e.g., knowledge of the results from using various amounts of detergent). Further, consumers need to be able to retrieve such information at the appropriate time. Although few studies have explored how memory for feedback influences judgment and how consumers weigh and integrate these various types of information, researchers have explored the effect of feedback about quantity decisions on usage. Several studies have addressed the effects of providing various types of information about energy use on conservation and about the amount of food ingested on dieting.

Feedback About Usage

Research on food consumption indicates that type of feedback is an important influence on consumption. Sometimes the type of feedback conducive to making quantity judgments is different from what might be expected. Contrary to everyday notions, physiological cues seem to be insufficient to enable people to specify the amount they ingested at a meal (Jordan 1969). The amount eaten by an adult at a single meal is not just physiologically determined by stomach capacity or

sensations of fullness, but is also psychologically influenced so that eating sometimes continues beyond sensations of fullness and sometimes ceases before capacity is reached (Booth 1994). In addition, the rate of consumption can also influence capacity, either increasing or decreasing the amount ingested. For example, fast eaters tend to overeat because physiological feedback from stomach fullness is delayed by about 20 minutes.

Consumption experience probably accounts for the important role of visual cues in influencing how much people eat. When not given visual cues about the amount of food eaten, even undistracted eaters are poor judges of the amount they have consumed. In addition to simple cues, calorie labels on foods help reduce the amount consumed (Kirschenbaum and Tomarken 1982). Multiple cues of different types may enhance the effectiveness of feedback because they increase its credibility and are likely to attract attention or facilitate recall of the information. People can be helped to monitor consumption if they are reminded to do so and monitoring is made salient. In one study, women who were easily able to monitor the amount of candy they had eaten by counting candy wrappers and women who were encouraged to count the number of cookies eaten ate less than those who could not monitor or were not encouraged to monitor the amounts (Polivy et al. 1986). Behavioral treatments for eating disorders commonly include regimens that encourage monitoring of the amount ingested (Stunkard and Mahoney 1974).

Research on energy consumption also supports the notion that feedback must be in a form that consumers can easily comprehend and relate to their own actions. When information about energy consumption is immediately available and detailed, people conserve more. Feedback on individual as opposed to group energy usage promotes conservation (Kasulis, Huettner, and Dikeman 1981), as does feedback on a daily versus a monthly basis (Hutton et al. 1986; Van Houwelingen and Van Raaij 1989).

The preceding points about the form of feedback may seem obvious, but feedback about quantities consumed is often difficult to comprehend and relate to one's behavior. The utility meter outside most homes provides immediate feedback, but not in a form most homeowners understand. Consumers conceive of energy consumption in terms of dollars and respond best when feedback is given in such terms (Kempton and Montgomery 1982). However, feedback in terms of dollars is misleading when inflation raises energy costs despite reductions in use. The household's conservation is not reinforced in the expected way when the bill arrives.

The literature on decision making in general supports the notion that feedback about the consequences of one's actions improves judgment. Yates (1990) suggests that detailed feedback about such factors as discrimination and slope of effects would be likely to improve judgment even more. Yet increasing the amount of information generally calls for a corresponding increase in the consumer's effort to process and use that information.

In sum, the amount and quality of feedback as well as the consumer's ability and motivation to use feedback influence the amount consumed. Most studies show that feedback helps consumers reduce consumption but that directional influence may be due to the products typically examined (food and energy). Depending on consumers' goals when using the product, feedback might also increase consumption. Knowledge of quantities can make salient any discrepancies between amounts consumed and goals sought. Feedback can also stimulate consumers to formulate targeted amounts before usage so that intentions control behavior before the onset of action (cf. Heckhausen and Beckmann 1990).

Important Issues in the Postactional Phase

More research is needed on motivational or directional biases in the search for, interpretation of, and use of consumption quantity feedback. "Garbology" studies in which actual usage was moni-

tored by analyzing refuse have shown that consumers underreport their own alcohol intake by 40 to 60 percent, whereas nondrinkers in the same household gauge others' alcohol intake with only about a 10 percent error (Rathje 1992). Comparing survey results with evidence from discards reveals that people overestimate the volume of healthy food they eat (e.g., liver, cottage cheese, tuna) and underestimate the volume of unhealthy food (e.g., sugar, candy, bacon) (Hudson 1984; Rathje 1992). The self-reports may reflect self-presentational strategies but also sincerely held beliefs based on biased search, interpretation, and use of feedback.

Another unexplored issue is related to whether the source of feedback is from usage or disposal. A consumer's product evaluation, and therefore usage, may be influenced by the act of disposing of a product. Discarding what are perceived as surplus amounts of a product may lead to a diminished evaluation of it. Disposing of a product may also interrupt the repetitive sequence of usage and prompt evaluation. Factors influencing the amount consumed should be examined across stages of consumer behavior (purchase, consumption, disposal).

Methodological Issues Pertinent to the Action-Goal Model

The action-goal model presented here organizes factors influencing usage decisions sequentially into one of four phases: the predecisional phase, the preactional phase, the execution phase, and the postactional phase. An alternative way to classify variables is into those that relate to the product (what is consumed), the person (who consumes), temporal factors (when it is consumed), and the locus of consumption (where it is consumed). Table 2.2 illustrates how some of the variables described in our chapter could be classified using the latter framework. The action-goal model seems a superior conceptual framework in its focus on the role of goals in motivating and regulating usage behavior. Nevertheless, the sequential nature of the model brings to the fore certain methodological issues that might not be as problematic for other frameworks. A particular challenge is identifying the beginning and termination of each phase. Additionally, the model's focus on the discrepancy between goals set for usage and factors causing consumers to deviate from those goals leads the researcher to anticipate differences between a consumer's cognitions about usage and actual behavior and to value measures of both. Accurate measures of the amount consumed require unobtrusive observation of the amount an individual uses during a single occasion.

Identifying Single Usage Occasions

Determining the unit of analysis when measuring the quantity consumed can be problematic when products are not used within narrow contexts. For example, measuring the amount of toothpaste applied on a single usage occasion is easy because usage is compressed into a short time period and occurs at regular intervals. Such patterns are less distinct for other products where diversity in the situation eliciting usage and temporal patterns vary. For example, a bottle of water in a person's car may be consumed at irregular intervals, so identifying the termination of the usage occasion is difficult for the researcher.

Collecting Usage Data

Another potential difficulty in usage quantity research is related to the measures used. Among the numerous obstacles to usage observation are privacy concerns. Observation of behavior is more feasible in public settings than private settings, but social influences on usage are more likely in public settings, which must be taken into account when drawing conclusions.

Table 2.2

Examples of Product, Person, Temporal, and Locus Factors Influencing Usage Quantity

Product	**Person**
Price	Knowledge
Instruction labels and warnings	Ownership of the product
Packaging	Materialism
Inventory size and unit size	Impulse control
Temporal	**Locus**
Time pressure	Presence of others
Continuous episodes	Salience of alternatives
Contiguous with feedback	Atmospherics
Contiguous with purchase	Variety of other products

Alternatively, the researcher might ask for recall of amounts used or predictions of amounts one would use. Many surveys include items designed to identify the amounts of products consumed by respondents. However, as noted in the review of research on food intake and energy consumption, people often lack knowledge about the amounts they have consumed. Instead of retrieving specific usage instances, an individual may rely on general knowledge to provide a response (Smith and Jobe 1994).

Further, recall of amounts used is hampered when measurement units are subjective. For example, the amount of food constituting a large portion to one person is only a moderate portion to another. Additionally, some populations may have different norms for portion sizes. College students define standard alcoholic drink volumes as being larger than the amounts commonly used by researchers (White et al. 2003). The overestimations varied by type of product. When students poured alcoholic beverages into a variety of cup sizes, they overestimated the amount they should pour for a serving size by 26 percent for shots, 80 percent for mixed drinks, and 25 percent for beer.

Consumption quantity researchers can sometimes estimate usage amounts by knowing the total quantity the individual consumed and the frequency of use, especially when the amount used is believed to be constant across occasions. In some cases, usage frequency data are sufficient to identify high-quantity users. Frequency and quantity are often correlated simply because the same factors can influence both. People who enjoy using a product will purchase more and have a larger supply on hand, which is likely to increase both the frequency and quantity of usage. However, frequency and single usage amounts can differ in their antecedents and consequences. Although correlated, frequency and quantity of drug use relates differentially to problem behaviors (Stein, Newcomb, and Bentler 1988).

Clearly, difficulties in obtaining precise measurements of quantities consumed can hamper research. Purchase quantity, in contrast, is relatively easy to monitor unobtrusively. However, unless consumption is contiguous with purchase, purchase quantity provides insight only into supplies available to the user.

Implications of Usage Research for Marketing and Public Policy

Our conclusions about the usage literature have implications for interventions by public policy makers and marketers. Although many of the behaviors described here may seem trivial and mundane, usage

quantity issues can have serious implications for consumers' safety and well-being. One study found that many physicians treating pulmonary tuberculosis prescribe insufficient amounts of drugs, thus imposing considerable financial costs for salvaging therapy (medical regimens that compensate for previous mistreatment) and causing detriment to the patient's health (Mahmoudi and Iseman 1993). Increasing the quantity of drugs used in that case would yield incalculable benefits.

Usage quantity has become increasingly linked with social policy concerns, which in turn attracts the attention of regulators and consequently can affect marketers. For example, the federal government's 1995 Dietary Guidelines for Americans suggest health benefits from drinking moderate amounts of alcohol, restricting consumption of frozen convenience foods and high-fat processed meats, increasing the amount of vegetables eaten, and exercise (*New York Times* 1996). Those guidelines influence school lunch programs, nutritionists' advice, and the food industry. Environmentalists' emphasis on reducing consumption (one of their three R's) has had wide influence (e.g., effects on taxation policies, public utility rates, marketing communications, and product design).

Besides relating to social policy concerns, usage quantity is relevant for firms' profits. Firms may want to increase sales by increasing the quantity of their offerings that purchasers consume. Profits can sometimes be increased by selectively increasing the quantity consumed of some products and decreasing that of others (e.g., a restaurant might profit by increasing the amount of beverages customers drink but decreasing the amount of condiments they use). Marketers may want to influence the quantity consumed because of its link with customer satisfaction. If consumers use amounts that are perceived to yield benefits (which could involve either an increase or a decrease in amounts), satisfaction will increase with corresponding gains for the marketer.

As is clear from the literature reviewed, firms generally can influence usage most easily by modifying aspects of the product (for examples, see Table 2.2). Governments can sometimes affect quantities consumed by influencing product factors (e.g., taxation policies influence price), the usage setting (e.g., restricted pub hours designed to decrease alcohol consumption), and a person's ability to use products (e.g., restricting teenagers' access to tobacco products). However, product factors may interact with other factors in the usage situation, so intervention may not be as simple as it seems. For example, increasing the amount prescribed for use on package instructions has the intended effect of increasing the amount the consumer believes should be used to gain the desired outcome. Yet when consumers have other sources of knowledge about product usage, they are likely to ignore instructions (Celuch, Lust, and Showers 1992). Previous government regulations restricting pub hours in England were intended to decrease the quantity of alcohol consumed. Those regulations were changed because they seemed to increase the amount consumed just prior to closing, leading to more social problems rather than fewer. Affecting the amount consumed requires thinking beyond simple product or situational modifications.

In sum, whereas many organizations have an interest in influencing consumers' usage quantity, doing so may be difficult because of lack of control over how consumers use products. Further, sometimes baseline consumption levels are present that are not easily changed. The amount of energy consumed, for example, is heavily influenced by house characteristics (Tienda and Aborampah 1981; Verhallen and Van Raaij 1981).

Nevertheless, even small changes in numerous individuals' behaviors can have large effects in the aggregate, especially for frequently consumed items. This point is readily acknowledged by those intent on the conservation of natural resources and is the basis for many ecologically sanctioned programs (e.g., recycling). The principle also is relevant for many consumer products. For example, small increases in the amount of catsup, dishwashing detergent, and toothpaste used can cumulate across individuals and occasions. A fast food firm increased its drinking straw size

and found that consumption of beverages increased. Apparently, customers reordered beverages because they finished their first drinks too quickly.

In short, even though consumption can be constrained by the capacity to consume, incremental changes can have an impact. The selection of a particular strategy to influence usage is likely to be a function of numerous factors, including the cost of implementing the strategy, the malleability of the behavior, and the magnitude of the change. However, costs can be influenced by the durability of the behavioral intervention (e.g., creating product variety might have only a temporary influence), the immediacy of the change, the compatibility of the usage message with other marketing efforts (e.g., with pricing strategies and advertising intended to influence purchase), and generality across consumers who have different decision-making strategies for usage.

Conclusions

Although many scholars have called for more attention to consumer behaviors other than product purchase, consumption has remained under-researched in the marketing literature. Among the aspects of consumption that warrant investigation is the quantity consumed. The relatively few marketing studies examining usage quantity issues may be due partly to the lack of a framework for integrating previous research on apparently disparate consumption quantity topics. Our chapter addresses that deficiency by providing an integrative framework for understanding usage quantity. Gollwitzer's (1996) action-goal model offers a means of understanding how various factors influence the quantity of consumption and identifying unresolved issues related to the quantity of consumption. Further, the model indicates the ways in which some factors may precede and therefore influence the impact of other factors.

Acknowledgments

The authors gratefully acknowledge the helpful comments of Michael Kamins, Deborah MacInnis, and C.W. Park on an earlier draft of this manuscript.

References

Aaronson, Lauren S. 1989. "Perceived and Received Support: Effects on Health Behavior During Pregnancy." *Nursing Research* 38: 4–9.

Ailawadi, Kusum, and Scott Neslin. 1998. "The Effect of Promotion on Consumption: Buying More and Using It Faster." *Journal of Marketing Research* 35: 390–398.

Anderson, C. Dennis, and John D. Claxton 1982. "Barriers to Consumer Choice of Energy Efficient Products." *Journal of Consumer Research* 9: 163–170.

Bagozzi, Richard P., and Elizabeth A. Edwards. 1998. "Goal Setting and Goal Pursuit in the Regulation of Body Weight." *Psychology and Health* 13: 593–621.

Bandura, Albert. 1977. "Self-Efficacy: Toward a Unifying Theory of Behavioral Change." *Psychological Review* 84: 191–215.

Basow, Susan A., and Diane Kobrynowicz. 1993. "What Is She Eating? The Effects of Meal Size on Impressions of a Female Eater." *Sex Roles* 28: 335–344.

Beggan, James K. 1992. "On the Social Nature of Nonsocial Perception: The Mere Ownership Effect." *Journal of Personality and Social Psychology* 62: 229–237.

Belk, Russell W. 1988. "Possessions and the Extended Self." *Journal of Consumer Research* 15: 139–168.

Bell, R., and Patricia L. Pliner. 2003. "Time to Eat: The Relationship Between the Number of People Eating and Meal Duration in Three Lunch Settings." *Appetite* 41: 215–218.

Bellisle, France, and Anne-Marie Dalix. 2001. "Cognitive Restraint Can Be Offset by Distraction, Leading to Increased Meal Intake in Women." *American Journal of Clinical Nutrition* 17: 197–200.

Bensley, Lillian. 1991. "Construct Validity Evidence for the Interpretation of Drinking Restraint as a Response Conflict." *Addictive Behaviors* 16 (3–4): 139–150.
Bettman, James R., and Mita Sujan. 1987. "Effects of Framing on Evaluation of Comparable and Noncomparable Alternatives by Expert and Novice Consumers." *Journal of Consumer Research* 14 (September): 141–154.
Bloch, Peter. 1995. "Seeking the Ideal Form: Product Design and Consumer Response." *Journal of Marketing* 59 (July): 16–29.
Booth, David A. 1994. *Psychology of Nutrition*. London: Taylor and Francis.
Breznitz, Shlomo. 1984. *Cry Wolf: The Psychology of False Alarms.* Hillsdale, NJ: Lawrence Erlbaum.
Caldwell, Clare, and Sally Hibbert. 2002. "The Influence of Music Tempo and Musical Preference on Restaurant Patrons' Behavior." *Psychology and Marketing* 19 (11): 895–917.
Cantor, Michael B., Stephen E. Smith, and Bonita R. Bryan. 1982. "Induced Bad Habits: Adjunctive Ingestion and Grooming in Human Subjects." *Appetite* 3: 1–12.
Celuch, Kevin G., John A. Lust, and Linda S. Showers. 1992. "Product Owner Manuals: An Exploratory Study of Nonreaders Versus Readers." *Journal of Applied Social Psychology* 22 (March): 492–507.
Chaiken, Shelly, and Patricia Pliner. 1987. "Women, But Not Men, Are What They Eat: The Effect of Meal Size and Gender on Perceived Femininity and Masculinity." *Personality and Social Psychology Bulletin* 13 (June): 166–176.
Chandon, Pierre, and Brian Wansink. 2002. "When Are Stockpiled Products Consumed Faster? A Convenience-Salience Framework of Post-Purchase Consumption Incidence and Quantity." *Journal of Marketing Research* 39 (August): 321–335.
Cornell, Carol E., Judith Rodin, and Harvey P. Weingarten. 1989. "Stimulus-Induced Eating When Satiated." *Physiology and Behavior* 45: 695–704.
Craig, C. Samuel, and John M. McCann. 1978. "Assessing Communication Effects on Energy Conservation." *Journal of Consumer Research* 5 (September): 82–88.
de Castro, John M. 1994. "Family and Friends Produce Greater Social Facilitation of Food Intake Than Other Companions." *Physiology and Behavior* 56: 445–455.
de Castro, John M., and E. Marie Brewer. 1992. "The Amount Eaten in Meals by Humans Is a Power Function of the Number of People Present." *Physiology and Behavior* 51: 121–125.
Deci, Edward L., and Richard M. Ryan. 1987. "The Support of Autonomy and the Control of Behavior." *Journal of Personality and Social Psychology* 53 (December): 1024–1037.
Dhar, Ravi, and Itamar Simonson. 1999. "Making Complementary Choices in Consumption Episodes: Highlighting Versus Balancing." *Journal of Marketing Research* 36 (February): 29–44.
Dickson, Peter R., and Alan G. Sawyer. 1990. "The Price Knowledge and Search of Supermarket Shoppers." *Journal of Marketing* 54 (July): 42–53.
Dunbar-Jacob, Jacqueline, Kathleen Dwyer, and E. Jean Dunning. 1991. "Compliance with Antihypertensive Regimen: A Review of the Research in the 1980s." *Annals of Behavioral Medicine* 13 (2): 31–39.
Edelman, Barbara, Diane Engell, Paul Gronstein, and Edward Hirsh. 1986. "Environmental Effects on the Intake of Overweight and Normal-Weight Men." *Appetite* 7 (March): 71–83.
Engell, Diane B., Morton Kramer, Teresa Malafi, Melinda Salomon, and Larry Lesher. 1996. "Effects of Effort and Social Modeling on Drinking in Humans." *Appetite* 26: 129–138.
Eisenberger, Robert. 1992. "Learned Industriousness." *Psychological Review* 99 (2): 248–267.
Fairburn, Christopher. 1993. *Binge Eating*. London: Guilford Press.
Folkes, Valerie S., Ingrid M. Martin, and Kamal Gupta. 1993. "When to Say When: Effects of Supply on Usage." *Journal of Consumer Research* 20 (December): 467–477.
Folkes, Valerie, and Shashi Matta. 2004. "The Effect of Package Shape on Consumers' Judgments of Product Volume: Attention as a Mental Contaminant." *Journal of Consumer Research* 31 (2): 390–401.
Fry, Prem S. 1975. "Affect and Resistance to Temptation." *Developmental Psychology* 11 (4): 466–472.
Gollwitzer, Peter M. 1996. "The Volitional Benefits of Planning." In *The Psychology of Action*, ed. Peter M. Gollwitzer and John A. Bargh. New York: Guilford Press.
Grube, Joel W., Genevieve M. Ames, and William Delaney. 1994. "Alcohol Expectancies and Workplace Drinking." *Journal of Applied Social Psychology* 24 (April): 646–660.
Hankin, Janet R., Ira J. Firestone, James J. Sloan, and Joel W. Ager. 1996. "Heeding the Alcoholic Beverage Warning Label During Pregnancy: Multiparae Versus Nulliparae." *Journal of Studies on Alcohol* 57 (March): 171–177.

Hearn M.D., Tom Baranowski, Janice Baranowski, Colleen Doyle, Lillian S. Lin, Matthew Smith, Donqing Terry Wang, and Ken Resnicow. 1998. "Environmental Determinants of Behavior Among Children: Availability and Accessibility of Fruits and Vegetables." *Journal of Health Education* 29: 26–32.

Heckhausen, Heinz, and Jurgen Beckmann. 1990. "Intentional Action and Action Slips." *Psychological Review* 97 (January): 36–48.

Helgeson, James G., E. Alan Kluge, John Mager, and Cheri Taylor. 1984. "Trends in Consumer Behavior Literature: A Content Analysis." *Journal of Consumer Research* 10 (March): 449–454.

Hoch, Stephen J., and John Deighton. 1989. "Managing What Consumers Learn from Experience." *Journal of Marketing* 53 (April): 1–20.

Holbrook, Morris B., and Elizabeth C. Hirschman. 1982. "The Experiential Aspects of Consumption: Consumer Fantasies, Feelings and Fun." *Journal of Consumer Research* 9 (September): 132–140.

Hudson, Luanne B. 1984. "Modern Material Culture Studies." *American Behavioral Scientist* 28 (September/October): 31–39.

Hutton, R. Bruce, Gary A. Mauser, Pierre Filiatrault, and Olli T. Ahtola. 1986. "Effects of Cost Related Feedback on Consumer Knowledge and Consumption Behavior: A Field Experimental Approach." *Journal of Consumer Research* 13 (December): 327–336.

Hutton, R. Bruce, and William L. Wilkie. 1980. "Life Cycle Cost: A New Form of Consumer Information." *Journal of Consumer Research* 6 (March): 349–360.

Jason, Leonard A., and Richard F. Liotta. 1982. "Reduction of Cigarette Smoking in a University Cafeteria." *Journal of Applied Behavior Analysis* 15 (Winter): 573–577.

Johnson, Michael. 1989."The Differential Processing of Product Category and Noncomparable Choice Alternatives." *Journal of Consumer Research* 16 (December): 300–309.

Jordan, Henry A. 1969. "Voluntary Intragastric Feeding: Oral and Gastric Contributions to Food Intake." *Journal of Comparative and Physiological Psychology* 68 (August): 498–506.

Kahn, Barbara E., and Alice M. Isen. 1993. "The Influence of Positive Affect on Variety Seeking Among Safe, Enjoyable Products." *Journal of Consumer Research* 20 (September): 257–270.

Kahn, Barbara E., and Brian Wansink. 2004. "The Influence of Assortment Structure on Perceived Variety and Consumption Quantities." *Journal of Consumer Research* 30 (4) (March): 519–533.

Kahneman, Daniel. 1992. "Reference Points, Anchors, Norms and Mixed Feelings." *Organizational Behavior and Human Decision Processes* 51 (March): 296–312.

Kamins, Michael A., and Henry Assael. 1987. "Two-Sided Versus One-Sided Appeals: A Cognitive Perspective on Argumentation, Source Derogation, and the Effect of Disconfirming Trial on Belief Change." *Journal of Marketing Research* 24 (February): 29–39.

Kassarjian, Harold H. 1978. "Anthropomorphism and Parsimony." In *Advances in Consumer Research* 5, ed. H. Keith Hunt, xii–xiv. Ann Arbor, MI: Association for Consumer Research.

Kasulis, Jack J., David A. Huettner, and Neil J. Dikeman. 1981. "The Feasibility of Changing Electricity Consumption Patterns." *Journal of Consumer Research* 8 (December): 279–290.

Kempton, Wittlet, and L. Montgomery. 1982. "Folk Quantification of Energy." *Energy* 7: 817–827.

Kirmani, Amna, and Peter Wright. 1989. "Money Talks: Perceived Advertising Expense and Expected Product Quality." *Journal of Consumer Research* 16 (December): 344–353.

Kirschenbaum, Daniel S., and Andres J. Tomarken. 1982. "Some Antecedents of Regulatory Eating by Restrained and Unrestrained Eaters." *Journal of Abnormal Psychology* 91 (October): 326–336.

Knaust, Gretchen, and Irene M. Foster. 2000. "Estimation of Food Guide Pyramid Serving Sizes by College Students." *Family and Consumer Sciences Research Journal* 29 (2): 101–110.

Knight, Linda J., and Fred J. Boland. 1989. "Restrained Eating: An Experimental Disentanglement of the Disinhibiting Variables of Perceived Calories and Food Type." *Journal of Abnormal Psychology* 98 (November): 412–420.

Linville, Patricia W., and Gregory W. Fischer. 1991. "Preferences for Separating or Combining Events." *Journal of Personality and Social Psychology* 60 (January): 5–23.

Loewenstein, George. 1994. "The Psychology of Curiosity." *Psychological Bulletin* 116 (July): 75–98.

Loewenstein, George, and Drazen Prelec. 1993. "Preferences for Sequences of Outcomes." *Psychological Review* 100 (January): 91–108.

Lyman, Bernard. 1989. *A Psychology of Food: More Than a Matter of Taste*. New York: Van Nostrand-Reinhold.

Mahmoudi, Artin, and Michael D. Iseman. 1993. "Pitfalls in the Care of Patients with Tuberculosis." *Journal of the American Medical Association* 270 (July 7): 65–68.

Martin, Ingrid, and Valerie S. Folkes. 2001. "Product Instructions as a Means of Fulfilling Consumer Usage Goals." In *European Advances in Consumer Research*, ed. Andrea Goppel-Klein and Franz-Rudolf Esch, 215–220. Valdosta, GA: Association for Consumer Research.

Milliman, Ronald E. 1986. "The Influence of Background Music on the Behavior of Restaurant Patrons." *Journal of Consumer Research* 13 (September): 286–289.

Mori, DeAnna, Shelly Chaiken, and Patricia Pliner. 1987. "'Eating Lightly' and the Self-Presentation of Femininity." *Journal of Personality and Social Psychology* 53 (October): 693–702.

Mustard, Tonya R., and Alastair V.E. Harris. 1989. "Problems in Understanding Prescription Labels." *Perceptual and Motor Skills* 69 (August): 291–294.

Nederkoorn, Chantal, and Anita Jansen. 2002. "Cue Reactivity and Regulation of Food Intake." *Eating Behaviors* 3: 61–72.

New York Times. 1996. "In About-Face, U.S. Sees Health Benefits of Alcohol." January 3, B6.

Norman, Donald A. 1988. *The Psychology of Everyday Things*. New York: Basic Books.

North, Adrian C., and David J. Hargreaves. 1996. "Responses to Music in a Dining Area." *Journal of Applied Social Psychology* 24: 491–501.

———. 1998. "The Effect of Music on Atmosphere and Purchase Intentions in a Cafeteria." *Journal of Applied Psychology* 28: 2254–2273.

O'Guinn, Thomas C., and Ronald Faber. 1989. "Compulsive Buying: A Phenomenological Exploration." *Journal of Consumer Research* 16 (September) 147–157.

Ohlsson, Stellan. 1996. "Learning from Performance Errors." *Psychological Review* 103 (April): 241–262.

Painter, James E., Bryan Wansink, and Julie Hieggelke. 2002. "How Visibility and Convenience Influence Candy Consumption." *Appetite* 38: 237–238.

Pan, Yigang, and Donald R. Lehmann. 1993. "The Influence of New Brand Entry on Subjective Brand Judgments." *Journal of Consumer Research* 20 (June): 76–86.

Parker, Louise E., and Mark R. Lepper. 1992. "Effects of Fantasy Contexts on Children's Learning and Motivation: Making Learning More Fun." *Journal of Personality and Social Psychology* 62 (April): 625–633.

Pennebaker, James W., and David Watson. 1988. "Blood Pressure Estimation and Beliefs among Normatensives and Hypertensives." *Health Psychology* 7 (4): 309–308.

Pliner, Patricia, and Shelly Chaiken. 1990. "Eating, Social Motives, and Self-Presentation in Women and Men." *Journal of Experimental Social Psychology* 3 (May): 240–254.

Polivy, Janet, C. Peter Herman, Rick Hackett, and Ira Kuleshnyk. 1986. "The Effects of Self-Attention and Public Attention on Eating in Restrained and Unrestrained Subjects." *Journal of Personality and Social Psychology* 50 (June): 1253–1260.

Poothullil, John M. 2002. "Role of Oral Sensory Signals in Determining Meal Size in Lean Women." *Nutrition* 18: 479–483.

Raghubir, Priya, and Aradhna Krishna. 1999. "Vital Dimensions in Volume Perception: Can the Eye Fool the Stomach?" *Journal of Marketing Research* 36 (August): 313–326.

Ragneskog, Hans, G. Bråne, I. Karlsson, and M. Kihlgren. 1996. "Influence of Dinner Music on Food Intake and Symptoms Common in Dementia." *Scandinavian Journal of Caring Sciences* 10 (1): 11–7.

Rathje, William L. 1992. *Rubbish! The Archeology of Garbage*. New York: HarperCollins.

Ratneshwar, Srinivasan, Allan D. Shocker, and David W. Stewart. 1987. "Toward Understanding the Attraction Effect: The Implications of Product Stimulus Meaningfulness and Familiarity." *Journal of Consumer Research* 13 (March): 520–533.

Richins, Marsha L. 1994. "Valuing Things: The Public and Private Meanings of Possessions." *Journal of Consumer Research* 21 (December): 504–521.

Richins, Marsha L., and Peter H. Bloch. 1986. "After the New Wears Off: The Temporal Context of Product Involvement." *Journal of Consumer Research* 13 (September): 280–285.

Richins, Marsha L., and Scott Dawson. 1992. "A Consumer Values Orientation for Materialism and Its Measurement: Scale Development and Validation." *Journal of Consumer Research* 19 (December): 303–316.

Rolls, Barbara J., Breda Kingston, Angela Megson, and Rachel Gunary. 1981. "Variety in a Meal Enhances Food Intake in Man." *Physiology and Behavior* 26 (February): 215–221.

Rolls, Edmund T., and J.H. Rolls. 1997. "Olfactory Sensory-Specific Satiety in Humans." *Physiology and Behavior* 61 (3): 461–473.

Rolls, Barbara J., Edward A. Rowe, and Edmund T. Rolls. 1982. "How Sensory Properties of Foods Affect Human Feeding Behavior." *Physiology and Behavior* 29 (February): 409–417.

Rolls, Barbara J., P.M. van Duijvenvoorde, and Edmund T. Rolls 1984. "Pleasantness Changes and Food Intake in a Varied Four-Course Meal." *Appetite* 5: 337–348.
Rook, Dennis W. 1985. "The Ritual Dimension of Consumer Behavior." *Journal of Consumer Research* 12 (December): 251–264.
Rose, Geoffrey A., and Richard Tecwyn Williams. 1961. "Metabolic Studies on Large and Small Eaters." *British Journal of Nutrition* 15 (1): 1–9.
Sadalla, Edward K., and Jennifer Krull. 1995. "Self-Presentational Barriers to Resource Conservation." *Environment and Behavior* 27 (May): 328–353.
Sansone, Carol, Charlene Weir, Lora Harpster, and Carolyn Morgan. 1992. "Once a Boring Task Always a Boring Task? Interest as a Self-Regulatory Mechanism." *Journal of Personality and Social Psychology* 63 (September): 379–390.
Shipman, Alan. 2004. "Lauding the Leisure Class: Symbolic Content and Conspicuous Consumption." *Review of Social Economy* 62 (3): 277–289.
Simonson, Itamar, and Russell Winer. 1992. "The Influence of Purchase Quantity and Display Format on Consumer Preference for Variety." *Journal of Consumer Research* 19 (June): 133–138.
Smith, Albert F., and Jared B. Jobe. 1994. "Validity of Reports of Long-Term Dietary Memories: Data and a Model." In *Autobiographical Memory and the Validity of Retrospective Reports*, ed. Norbert Schwarz and Seymour Sudman, 121–140. New York: Springer-Verlag.
Smith, Robert E. 1994. "Integrating Information from Advertising and Trial: Processes and Effects on Consumer Response to Product Information." *Journal of Marketing Research* 28 (May): 204–219.
Snyder, C.R. 1992. "Product Scarcity by Need for Uniqueness Interaction: A Consumer Catch-22 Carousel." *Basic and Applied Social Psychology* 13 (1): 9–24.
Stein, Judith A., Michael D. Newcomb, and P.M. Bentler. 1988. "Structure of Drug Use Behaviors and Consequences Among Young Adults: Multitrait-Multimethod Assessment of Frequency, Quantity, Work Site, and Problem Substance Use." *Journal of Applied Psychology* 73 (4): 595–605.
Stewart, David W., Valerie S. Folkes, and Ingrid M. Martin. 2001. "Consumer Response to Warnings and Other Types of Product Hazard Information: Future Public Policy and Research Directions." In *Handbook of Marketing and Society*, ed. Paul N. Bloom and Gregory T. Gundlach, 335–371. Thousand Oaks, CA: Sage.
Stroebele, Nanette, and John M. de Castro. 2004. "Television Viewing Is Associated with an Increase in Meal Frequency in Humans." *Appetite* 42: 111–113.
Stunkard, Albert J., and J.J. Mahoney. 1974. "Behavioral Treatment of the Eating Disorders." In *Handbook of Behavior Modification*, ed. Harold Leitenberg New York: Appleton-Century-Crofts.
Teahan, Joln. 1987. "Alcohol Expectancies, Values and Drinking of Irish and US Collegians." *International Journal of the Addictions* 22 (July): 621–638.
Thaler, Richard. 1985. "Mental Accounting and Consumer Choice." *Marketing Science* 4 (Summer): 199–214.
Thaler, Richard, and Eric Johnson. 1990. "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice." *Management Science* 36 (June): 643–660.
Tienda, Marta, and Osei-Mensah Aborampah. 1981. "Energy-Related Adaptations in Low-Income Nonmetropolitan Wisconsin Counties." *Journal of Consumer Research* 8 (3): 265–270.
Tucker, Larry A., and M.R. Bagwell. 1991. "Television Viewing and Obesity in Adult Females." *American Journal of Public Health* 81: 908–111.
Tucker, Larry A., and G.M. Friedman. 1989. "Television Viewing and Obesity in Adult Males." *American Journal of Public Health* 79: 516–518.
Tykocinski, Orit E., Thane S. Pittman, and Erin E. Tuttle. 1995. "Inaction Inertia: Foregoing Future Benefits as a Result of an Initial Failure to Act." *Journal of Personality and Social Psychology* 68 (5): 793–803.
Van Houwelingen, Jennet H., and W. Fred Van Raaij. 1989. "The Effect of Goal-Setting and Daily Electronic Feedback on In-Home Energy Use." *Journal of Consumer Research* 16 (June): 98–105.
Verhallen, Theo M.M., and W. Fred Van Raaij. 1981. "Household Behavior and the Use of Natural Gas for Home Heating." *Journal of Consumer Research* 8 (December): 253–257.
Wadhwa, Monica, Baba Shiv, and Stephen M. Nowlis. 2006. "A Bite to Whet the Reward Appetite: Influence of Sampling on Reward-Seeking Behaviors." Working paper, Stanford University.
Wansink, Brian. 1996. "Can Package Size Accelerate Usage Volume?" *Journal of Marketing* 60 (July): 1–14.
Wansink, Brian, and Pierre Chandon. 2006. "Can 'Low Fat' Nutrition Labels Lead to Obesity?" *Journal of Marketing Research* 43 (November), 605–617.

Wansink, Brian, and Michael L. Ray. 1996. "Advertising Strategies to Increase Usage Frequency." *Journal of Marketing* 60 (January): 31–46.

Wansink, Brian, and Koert van Ittersum. 2003. "Bottoms Up! The Influence of Elongation and Pouring on Consumption Volume." *Journal of Consumer Research* 30 (December): 455–463.

Weiner, Bernard. 1986. *An Attributional Theory of Motivation and Emotion*. New York: Springer-Verlag.

Wells, William D. 1993. "Discovery-Oriented Consumer Research." *Journal of Consumer Research* 19 (March): 489–504.

Wertenbroch, Klaus. 1998. "Consumption Self-Control by Rationing Purchase Quantities of Virtue and Vice." *Marketing Science* 17 (4): 317–337.

White, Aaron M., Courtney L. Kraus, Lindsey A. McCracken, and H. Scott Swartzwelder. 2003. "Do College Students Drink More Than They Think? Use of a Free-Pour Paradigm to Determine How College Students Define Standard Drinks. Neurobiological, Behavioral and Environmental Relations to Drinking." *Alcoholism: Clinical & Experimental Research* 27 (11): 1750–1756.

Winett, Richard A., and John H. Kagel. 1984. "Effects of Information Presentation Format on Resource Use in Field Studies." *Journal of Consumer Research* 11 (September): 655–667.

Wing, Rena R., and Roert W. Jeffery. 2001. "Food Provision as a Strategy to Promote Weight Loss." *Obesity Research* 9: 271S–275S.

Worchel, Stephen, Jerry Lee, and Akanbi Adewole. 1975. "Effects of Supply and Demand on Ratings of Object Value." *Journal of Personality and Social Psychology* 32 (November) 906–914.

Wright, Patricia. 1981. "The Instructions Clearly State. Can't People Read?" *Applied Ergonomics* 12 (September): 131–141.

Wright, Patricia, Patricia Creighton, and S.M. Threlfall. 1982. "Some Factors Determining When Instructions Will Be Read." *Ergonomics* 25 (March): 225–237.

Wright, Peter. 1986. "Schemer Schema: Consumers' Intuitive Theories About Marketers' Influence Tactics." In *Advances in Consumer Research* 13, ed. Richard J. Lutz, 1–3. Provo, UT: Association for Consumer Research.

Yates, J. Frank. 1990. *Judgment and Decision Making*. Englewood Cliffs, NJ: Prentice Hall.

CHAPTER 3

INTEGRATING PURCHASE TIMING, CHOICE, AND QUANTITY DECISIONS MODELS

A Review of Model Specifications, Estimations, and Applications

V. KUMAR AND ANITA MAN LUO

Abstract

Knowing what to sell, when to sell, and to whom is essential for a firm to allocate scarce marketing resources in an efficient and effective manner. This study reviews how purchase timing, brand choice, and purchase quantity decisions have been historically modeled and what issues within each decision have been addressed. For example, issues like cross-category dependence and state dependence have generated a tremendous amount of research interest. In addition, we examine the fundamental differences among various approaches and present the common methods that have been used to model at least two of the three decisions. Most of the studies we reviewed have acknowledged the dependence among the three decisions through various methodologies. Finally, we discuss the managerial applications of modeling these decisions and suggest ways to address the future challenges.

Introduction

Business executives today want to know when consumers make a purchase, and if they do, what brands/categories they choose and how much they buy. Understanding what to sell, when to sell, and to whom is essential for a firm to allocate scarce marketing resources in an efficient and effective manner.

Extensive research has been done in marketing to capture the timing, choice, and quantity decisions. Within the timing models, previous research has adopted the approach of *whether to buy* or *when to buy*, depending on the researchers' point of view. It is also widely believed that there is a cross-category dependence, which captures a consumer's tendency to buy products from certain categories together. Therefore, depending on a researcher's interest and modeling assumptions, the researcher has the liberty to model timing through one of the approaches and decide whether or not to address multicategory dependence.

A major topic that has captured researchers' interests in the brand choice model is state dependence. Academics as well as practitioners are interested to know how a consumer's previous

choice of a brand affects his or her future brand choice behavior. Previous research has addressed this issue through some measurements of brand loyalty, lagged choice, and so forth. A stream of research has also been dedicated exclusively to the investigation of state dependence.

Depending on the assumptions of the interdependence of the three decisions, purchase timing, brand choice, and purchase quantity have been modeled individually or jointly in various ways. Most of the literature we have reviewed has tried to address the interrelationships for reasons as shown below: (1) the modeling of the three decisions separately cannot account for the selection bias in purchase quantity and choice models; (2) the three decisions may not be independent given that the same mechanism such as income, marketing response elasticity, or consumption pattern may drive these three decision processes simultaneously; (3) the jointly estimated model allows a consumer to maximize bundle utility over a period of time and therefore reflects the dynamically changing behavior of a consumer. On the contrary, a separately estimated model views a consumer's purchase behavior as a snapshot.

Based on the input of a customer's behavior in terms of purchase timing, brand choice, and purchase quantity, a firm can decide what to sell, when, and to whom in order to maximize profitability at the individual customer level as well as at the firm level. Kumar, Venkatesan, and Reinartz (2006) suggested that by using data from the customer relationship management system, a firm could determine the probability that an existing customer will purchase a given product at a given time and thereby target the customers who are most likely to purchase. This prediction helps to customize the offer.

In the following sections, we will first review how the purchase timing, brand choice, and purchase quantity decisions have been historically modeled and what issues within each decision have been addressed. In addition, we will add a dimension to the discussion: whom to target based on the understanding of the three decisions. Then, we will review the fundamental differences among various approaches. We believe that most of the differences rely on the assumptions of a consumer's decision-making process and how unobserved heterogeneity can be captured. Next, we will present the common approaches that have been used to model at least two of the three decisions along with how these models are specified. Finally, we will discuss managerial applications of these decisions and suggestions for future research.

Review of Purchase Timing, Choice, and Quantity Models

In the past several decades, we have witnessed a large amount of literature dedicated to the modeling of purchase timing, choice, and quantity decisions. Table 3.1 lists some representative studies that have examined at least two of the three decisions. As you can see from Table 3.1, we also indicated if unobserved heterogeneity was captured, what type of data were used, and what the business and industry settings were for the study.

Questions that have been addressed in this stream of literature as shown in Figure 3.1 include: How are the three decisions interrelated? What are the different approaches that were used to model each of the three decisions? What are the roles that marketing variables and consumer heterogeneity played in a consumer's purchase decisions? Figure 3.1 provides an overview of issues that have generated research interests in this field. As shown in Figure 3.1, a consumer faces three decisions in a purchase occasion: whether to buy in a category, which brand to choose within that category, and how much to buy. A consumer's decision to purchase in a category depends on the timing of the previous purchase in the category, the decision to buy in a related category, the marketing variables, and consumer heterogeneity. Similarly, the choice of brand is determined by a consumer's previous choice of brand through state dependence, the relative utilities of brands in

Table 3.1

Summary of Select Empirical Studies Modeling at Least Two Decisions

Studies	Timing/ Incidence	Choice	Quantity	Unobserved Heterogeneity	Nature of Data	Business Type	Industry Type
Neslin, Henderson, and Quelch (1985)	Yes	—	Yes	No	Scanner data	B-C	Retailer
Krishnamurthi and Raj (1988)	—	Yes	Yes	No	Scanner data	B-C	Retailer
Tellis (1988)	—	Yes	Yes	No	Scanner data	B-C	Retailer
Gupta (1988)	Yes	Yes	Yes	No	Scanner data	B-C	Retailer
Bucklin and Lattin (1991)	Yes	Yes	—	No	Scanner data	B-C	Retailer
Chiang (1991)	Yes	Yes	Yes	No	Scanner data	B-C	Retailer
Bucklin and Gupta (1992)	Yes	Yes	—	Yes	Scanner data	B-C	Retailer
Chintagunta (1993)	Yes	Yes	Yes	Yes	Scanner data	B-C	Retailer
Ainslie and Rossi (1998)	Yes	Yes	—	Yes	Scanner data	B-C	Retailer
Arora, Allenby, and Ginter (1998)	Yes	Yes	Yes	Yes	Survey data	B-C	Retailer
Boatwright, Borle, and Kadane (2003)	Yes	—	Yes	Yes	Transaction data	B-C	Online retailer
Zhang and Krishnamurthi (2004)	Yes	Yes	Yes	Yes	Transaction data	B-C	Online retailer
Niraj, Padmanabhan, and Seetharaman (2007)	Yes	—	Yes	Yes	Scanner data	B-C	Retailer
Gönül and Ter Hofstede (2006)	Yes	—	Yes	Yes	Transaction data	B-C	Catalog
Song and Chintagunta (2006)	Yes	Yes	—	—	Aggregate store data	B-C	Retailer
Kumar, Venkatesan, and Reinartz (2006)	Yes	Yes	—	Yes	Transaction data	B-B	High-tech

Figure 3.1 **Issues Addressed in Purchase Timing, Choice, and Quantity Decision Models**

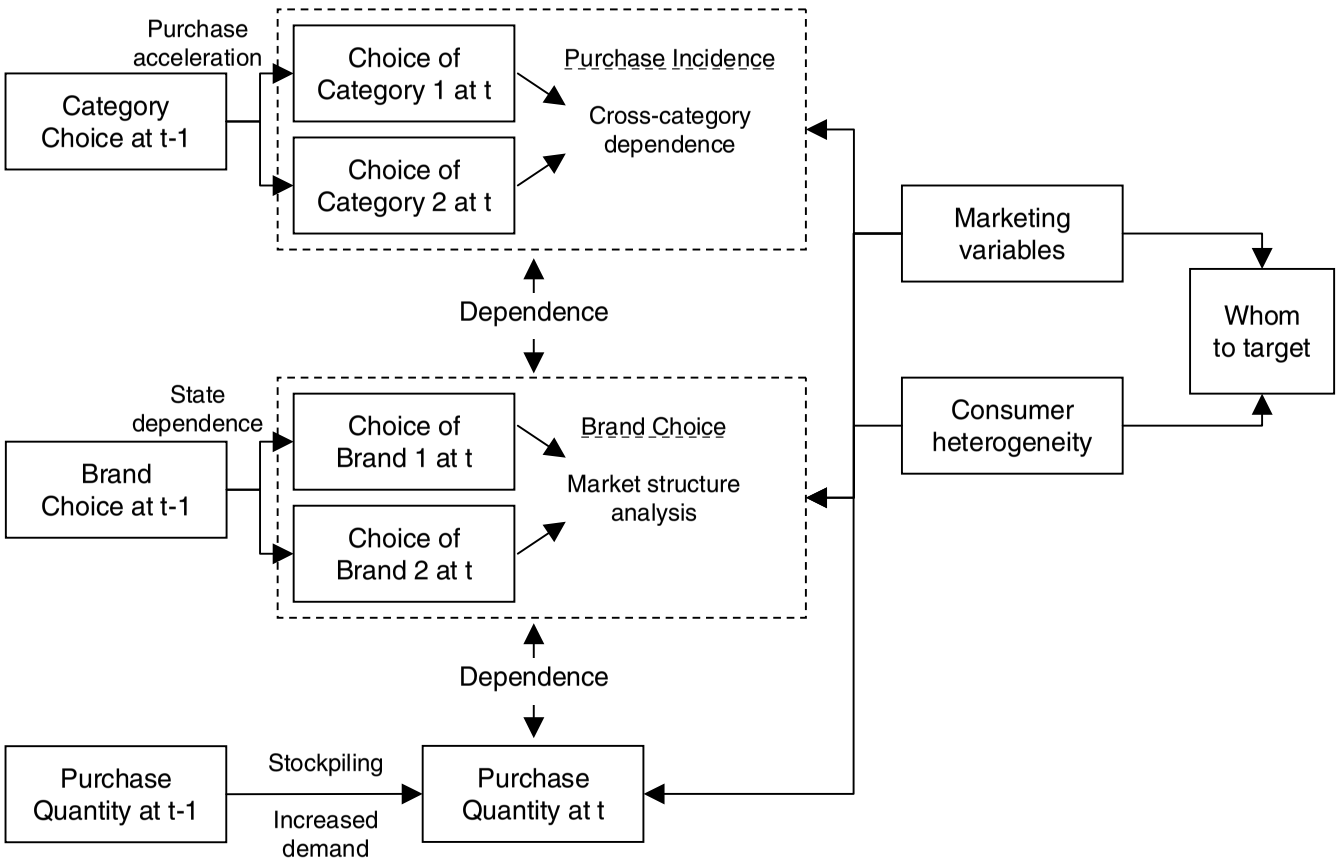


Table 3.2

Review of Purchase Timing/Incidence Related Models

Research Interest	Specification	Representative Studies
Whether to buy	Logit	Bucklin and Lattin (1991) Bucklin and Gupta (1992) Zhang and Krishnamurthi (2004) Niraj, Padmanabhan, and Seetharaman(2007)
	Single utility function	Chiang (1991) Chintagunta (1993) Arora, Allenby and Ginter (1998)
When to buy	Distribution of elapse time	Gupta (1988)—Erlang Distribution Boatwright, Borle, and Kadane (2003)—CM Poisson
	Hazard function	Jain and Vilcassim (1991) Seetharaman and Chintagunta (2003) Gönül and Ter Hofstede (2006) Kumar, Venkatesan, and Reinartz (2006)
Overview of cross-category dependence	—	Ainslie and Rossi (1998) Manchanda, Ansari, and Gupta (1999) Gentzkow (2006)

the consideration set, marketing activities, and the consumer's heterogeneous brand preference. Finally, a consumer's heterogeneous response to marketing variables and inventory level affect a consumer's purchase quantity decision. To provide a comprehensive understanding of how to capture the three decisions, we will first start with a historical review of how to model each of the decisions as described in the studies (Table 3.1) and also include a few studies that provide insights into modeling purchase timing, choice, and quantity alone.

Purchase Timing

In order to model purchase timing, one has to decide which approach is more suitable: *whether to buy* or *when to buy.* In this section, we will first go through the two schools of thought and then discuss studies that also examine cross-category dependence (Table 3.2). As shown in Table 3.2, the general approaches of modeling *whether to buy* are either through a logit function for purchase timing decisions alone or a single direct utility function incorporating purchase timing, brand choice, and quantity decisions. To estimate *when to buy*, one can propose a parametric distribution such as Erlang-2 for the elapse time or capture such phenomena within the framework of Hazard Function.

Whether to Buy

Wheat and Morrison (1990) suggested that *whether to buy* performs better than *when to buy* for the following two reasons: the purchase incidence model can well capture the discrete nature of purchase pattern; the inter-purchase time model is subject to the limitation of right censored data. Researchers like Chiang (1991) have modeled *whether to buy* rather than *when to buy.* Chiang (1991) suggested that the decision of *when to shop* governs the decision of *when to buy.* He observed that the histograms of the inter-purchase time for many product categories show a spike at

a seven-day interval. If most of the consumers adopt a considerably strict shopping routine, the purchase timing decision can be modeled through the conditional probability of *whether to buy* given each shopping occasion.

Chiang (1991) and Chintagunta (1993) modeled purchase incidence probability (*whether to buy*) by comparing the reservation price with the lowest adjusted price of the brands in a category. In their model, the adjusted price equals the market price divided by the quality index. Therefore, there are two components in the model that need to be correctly measured: the reservation price and the quality index of various brands in a product category. The quality index is a function of product attributes, marketing mix variables such as feature and display, and consumer-specific variables such as past purchasing experience. Chiang (1991) specified threshold price as a function of the price of composite goods divided by their product attributes and marketing variables, since a consumer allocates the shopping budget between the composite goods and the product of interest. Chintagunta (1993) assumed a parametric and semi-parametric form for the reservation price across consumers. In his model, consumer inventory level is also incorporated in the specification of the quality attributes to capture a consumer's intrinsic need for the product. Chintagunta suggested that store visits in which a consumer is exposed to the marketing variables but does not make a category purchase should also be incorporated in the analysis. If only marketing information at the time of purchase is incorporated, the effect of marketing variables might be overstated. Therefore, this type of model allows researchers to examine the effect of marketing information even when a consumer is not buying in the category.

Chiang (1991) and Chintagunta (1993) both assume that a consumer will evaluate various brands in a category and make a purchase incidence decision depending on marketing variables such as price promotion and the consumer's intrinsic reservation price for the product category. However, is it true that a consumer always compares the option of buying versus not buying in a category in each shopping trip? Bucklin and Lattin (1991) suggested that there are two types of shopping occasions: planned state and opportunistic state. The difference between the two states depends on whether a consumer has considered a purchase and made a decision to buy a given brand or not to buy at all (Bucklin and Lattin 1991). The probability of a consumer being in a planned or opportunistic state is determined by deal loyalty, inventory, and loyalty to the store. Given the shopping state a consumer is in, purchase incidence is modeled as a function of category consumption, derived inventory, and category value. In their model, marketing variables such as price and promotion influence purchase incidence through category value. The regularity of purchase pattern is captured through the category consumption rate and inventory. Bucklin and Gupta (1992) also modeled purchase incidence through a similar specification. They suggested that consumption rate and inventory can be used to capture heterogeneity across consumers over time. In this approach, they measured consumption rate as the total amount purchased by a consumer divided by the number of weeks in the initialization period. Consumer inventory at the current time period is determined by adding quantity purchased in the last period to the inventory of that period, and then deducting the consumption rate. Therefore, the consumption rate and inventory are deterministically computed without making adjustments for variations in measurement error. In addition, the consumption rate is assumed to be constant over time for each consumer and does not vary with changes in marketing conditions. This type of model thus assumes that a consumer makes sequential decisions of shopping occasion, purchase incidence, and brand choice. Given a shopping occasion, purchase incidence is determined by a consumer's intrinsic need for the product category and the overall attractiveness of brands in a product category.

Zhang and Krishnamurthi (2004) also modeled purchase incidence and brand choice through a logit model. The utility of making no purchase in a category is a function of a consumer's pur-

Table 3.3

Hazard Function and Survival Function of Various Baseline Distributions

Baseline Distribution	Hazard Function	Survival Function
Exponential	$h(t) = \gamma$	$S(t) = e^{-\gamma t}$
Erlang-2	$h(t) = \frac{\gamma^2 t}{1+\gamma t}$	$S(t) = (1+\gamma t)e^{-\gamma t}$
Weibull	$h(t) = \gamma\alpha(\gamma t)^{\alpha-1}$	$S(t) = e^{-(\gamma t)^{\alpha}}$
Log-logistic	$h(t) = \frac{\gamma\alpha(\gamma t)^{\alpha-1}}{1+(\gamma t)^{\alpha}}$	$S(t) = \frac{1}{1+(\gamma t)^{\alpha}}$
Expo-power	$h(t) = \gamma\alpha\, t^{\alpha-1} e^{\theta t^{\alpha}}$	$S(t) = e^{\frac{r}{\theta}[1-e^{\theta t^{\alpha}}]}$

Source: Seetharaman and Chintagunta (2003).

chase frequency in the initialization period and the mean-centered last purchase quantity before the current shopping trip. They did not use an inventory variable, since Chintagunta and Haldar (1998) suggested that such a variable may be endogenous to the purchase incidence decision. In conclusion, researchers can model *whether to buy* through the approaches discussed above.

When to Buy

As we mentioned before, purchase timing can also be modeled through *when to buy.* The estimation of *when to buy* usually starts with a parametric form of the inter-purchase time. The negative binomial distribution model proposed by Ehrenberg (1959) suggests that the inter-purchase time follows an exponential distribution. However, as suggested by the *memory-less property* of the exponential distribution, the fact that a consumer has not yet bought tells us nothing about how long it will take for the consumer to purchase. Therefore, such an assumption may not be valid in situations where the probability of buying changes as the time since previous purchase increases.

Because of the limitation imposed by the *memory-less property* of the exponential distribution, Gupta (1988) used Erlang-2 distribution to model inter-purchase time. Chatfield and Goodhardt (1973) suggested that an Erlang distribution could be viewed as a censored Poisson process counting only every *p*th event. However, Erlang distribution may not always perform well in modeling inter-purchase timing. Jain and Vilcassim (1991) stated that there is often no available theory to specify a probability distribution for the inter-purchase times, and one possible solution is to test a variety of distributions and select the one that is most suitable for the data. Seetharaman and Chintagunta (2003) proposed several baseline hazard specifications to capture the general purchase pattern: exponential, Erlang-2, Weibull, log-logistic, and expo-power. Table 3.3 lists the hazard functions and the survival functions of various baseline distributions as specified in their paper. It is essential to choose the appropriate baseline hazard depending on the purchase pattern in a product category. In general, a consumer's demand for durable goods may decrease as time since last purchase increases and then start to increase again once reaching a threshold. Therefore, an expo-power distribution may be suitable for such situation.

According to Seetharaman and Chintagunta (2003), the baseline hazard for Erlang-2 distribution is monotonically increasing as the time since last purchase increases (Figure 3.2), the hazard for Weibull distribution can be monotonically increasing or decreasing (Figure 3.3), and the hazard for log-logistics can be monotonically decreasing or inverted U-shaped (Figure 3.3). They suggested that expo-power performs the best given its flexibility in a parametric form, since it can be flat, monotonically increasing or decreasing, U-shaped, or inverted U-shaped (Figure 3.4). They also stated that the model estimation results may be quite sensitive to the chosen parametric form. Thus, the choice of the baseline hazard is critical and category dependent.

Jain and Vilcassim (1991) suggested that it is essential to adjust for the effect of marketing mix variables and unobserved heterogeneity to best capture the baseline function. They specified a flexible parametric form allowing for competing baseline hazard distributions. The marketing covariates can proportionally shift the baseline hazard rate up or down. Unobserved heterogeneity is captured with a distribution across consumers through a random coefficient model. Therefore, the multiple spells of a single consumer's purchase history won't be treated as spells from different consumers when a consumer-specific parameter is included in the model. The consequence of choosing a wrong baseline distribution results in poor prediction.

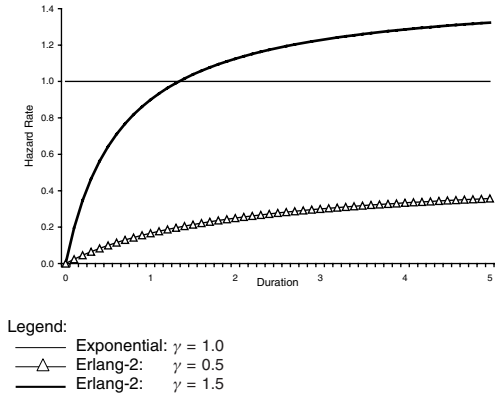
As already discussed, a continuous hazard model cannot capture discrete purchase behavior in an appropriate manner (Wheat and Morrison 1990). If a consumer has a strict shopping schedule, the purchase probability for the consumer in certain time periods may be zero. Therefore, a discrete proportional hazard model might be more appropriate if there are observed spikes in regular intervals of inter-purchase time.

Similar to the approach of *whether to buy*, a discrete hazard model can incorporate marketing information when no purchase of the product category is observed in a shopping occasion. Seetharaman and Chintagunta (2003) suggested that it is essential to incorporate information from the observed events of a consumer's nonpurchase in a product category despite the fact that the observed times of a consumer's purchase is the central interest. Finally, another advantage of the discrete hazard model is that it is not subjected to the limitation of right censored data as the continuous hazard model is.

Seetharaman and Chintagunta (2003) further developed two cause-specific hazards for situations where a consumer made a purchase during his previous shopping trip and where a consumer did not buy anything. Such a specification can also capture the effect of consumer inventory on purchase probability, since now the discrete hazard rate depends on whether the product was bought in the previous shopping trip. They also proposed a nonparametric proportional hazard model because the chosen parametric form for the baseline hazard may affect the estimation of the marketing covariates parameters. Therefore, a time-specific intercept-term that varies freely from one discrete time period to another, rather than a baseline function integrated over a discrete time period, is used. However, they stated that the disadvantage of such nonparametric specification is the increased number of parameters to be estimated (Seetharaman and Chintagunta 2003).

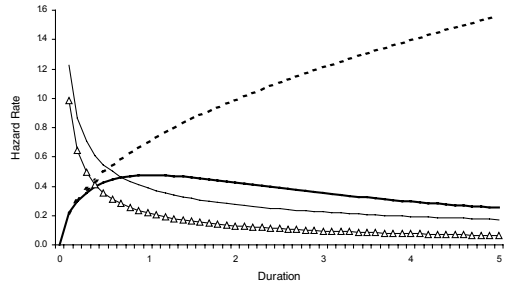
The proportional hazard models estimate the hazard rate based on the time since a consumer's previous purchase. Bijwaard (2006) commented that each inter-purchase time of a consumer in the traditional proportional hazard model is modeled separately without taking into consideration the consumer's purchase history. In addition, the proportional hazard function cannot predict whether a consumer will make two or more purchases in a certain period. Therefore, Bijwaard proposed that calendar time instead of elapse time should be used so that information since the observed beginning of a consumer's purchase history can be incorporated. He further decomposed the hazard into four components: the calendar duration dependence modeling seasonal effects,

Figure 3.2 **Hazard Functions for Exponential and Erlang-2**



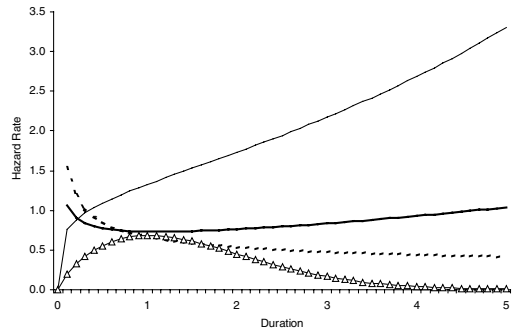
Legend:
Exponential: $\gamma = 1.0$
Erlang-2: $\gamma = 0.5$
Erlang-2: $\gamma = 1.5$

Figure 3.3 **Hazard Functions for Log-Logistic and Weibull**



Legend:
Log-Logistic: $\gamma = 0.6$ $\alpha = 1.5$
Log-Logistic: $\gamma = 0.6$ $\alpha = 0.5$
Weibull: $\gamma = 0.6$ $\alpha = 0.5$
Weibull: $\gamma = 0.6$ $\alpha = 1.5$

Figure 3.4 **Hazard Functions for Expo-Power**



Legend:
Expo-power: $\theta = 0.1,\ \gamma = 1,\ \alpha = 1.2$
Expo-power: $\theta = 0.1,\ \gamma = 1,\ \alpha = 0.6$
Expo-power: $\theta = -0.4,\ \gamma = 0.6,\ \alpha = 1.7$
Expo-power: $\theta = 0.4,\ \gamma = 0.7,\ \alpha = 0.7$

the gap duration dependence capturing the intrinsic purchase pattern, the covariate function, and the unobserved heterogeneity.

In his model, he also specified a risk indicator, which can take different values indicating whether a consumer is at the risk for making the second purchase, third purchase, and so forth. Bijwaard (2006) suggested that the at-risk indicator be set to zero after each purchase in the traditional hazard models, whereas the risk remains as long as the consumer has not left the study in his model. In addition, if a consumer is not likely to purchase in certain days of a week, the risk indicator can be set to zero for those days.

Another assumption that the proportional hazard model makes is that marketing variables have a multiplicative effect on the hazard rate. Seetharaman (2004) proposed additive risk model to capture purchase timing, since marketing variables may affect the hazard rate in an additive manner. The additive risk model is very similar to the proportional hazard model except that marketing effects are added to the baseline hazard rate. Seetharaman found out that the additive risk model outperforms the proportional hazard model in explaining purchases. However, such a finding needs further verification under different circumstances.

Unlike household panel data, certain data do not have full information about a consumer's purchase history. For example, companies in general only have records of consumers buying their own products. They lack information as to whether the consumers are buying from the competitors. Boatwright, Borle, and Kadane (2003) faced such problem when they tried to model online grocery purchases. If a consumer is buying exclusively from the online retailer, the authors argue that the purchase quantity should be roughly proportional to the elapsed time. Therefore, the dependence between purchase quantity and inter-purchase time through inventory control tells us if a consumer is buying from competitors or not when such information is not available in the data. However, the assumption here is that the consumer has a fairly stable demand for the product. Therefore, variation in inter-purchase time is explained either by stockpiling in the previous trip or by buying from competitors. However, when a customer has very flexible demand for a certain product such as durable goods, such specification could have a limited ability to correctly estimate the purchase pattern and the effect of marketing variables given only partial information.

Therefore, Kumar et al. (2006) used a data augmentation method to impute *missing data* when the time since last purchase is greater than the average inter-purchase time calculated in the calibration sample. They imputed the missing observations based on a normally distributed prior distribution with mean and variance empirically estimated from the calibration sample. In a nutshell, choosing a model to capture inter-purchase times depend on a researcher's view of a consumer's behavior and the problem at hand.

Cross-Category Dependence

Most of the models we discussed above focused only on a single-category purchase. However, consumers usually make multicategory purchases during each shopping trip. Due to unobserved heterogeneity such as reservation price and response to marketing variables, one shall question whether there is a similar purchase-timing pattern across multiple categories. Therefore, one might ask, does the probability of buying in one category depend on buying in another category? In other words, are two categories positively related, negatively related, or independent? If we observe that products from two categories are regularly purchased together in a shopping basket, what are the reasons behind the combined purchase?

Ainslie and Rossi (1998) examined similarity in choice behavior across product categories by measuring observed and unobserved heterogeneity in marketing mix sensitivity across multiple

categories. First, they demonstrated that the correlation between marketing sensitivity of two categories is always understated because of estimation errors. In order to rectify the bias, they suggested an approach modeling the purchase of products of all categories simultaneously. They decomposed the marketing mix response parameters into two parts: one is consumer-specific and does not vary across categories, and the other is category-specific. The consumer-specific component is a function of consumer characteristics such as demographics and overall shopping behavior and normal random error, while the category-specific component is a function of category-specific variables such as category indicator and normal random error. Therefore, the unconditional covariance matrices of two categories incorporate both observed and unobserved correlation in sensitivities. They used a Bayesian hierarchical approach to obtain the posterior distribution of marketing variable sensitivities through Gibbs sampling.

Seetharaman et al. (2005) suggested that there is a base level of complementarity estimated from all pairs of product categories, since a large number of no-purchase outcomes are observed across product categories. Ma and Seetharaman (2006) estimated cross-category correlations separately for purchase outcome and nonpurchase outcome through the multivariate logit model.

Manchanda, Ansari, and Gupta (1999, p. 98) suggested that complementarity, co-incidence, and heterogeneity could influence a consumer's choice to simultaneously buy in two categories. They defined two categories as complements "if marketing actions (price and promotion) in one category influence the purchase decision in the other category." They also defined co-incidence as "the set of all reasons except purchase complementarity and consumer heterogeneity that could induce joint purchase of items across categories." They specified the utility of buying in one category is a function of marketing variables of its own category and those of another category. Therefore, the effect of marketing variables of another category decides if two categories are complements, substitutes, or related depending on if it can increase, decrease, or has no effect on the purchase probability in the focal category. The random component of the utilities of buying in different categories follows a multivariate probit distribution. Therefore, the covariance of the random component of buying in different categories captures coincidence. They also used a Bayesian hierarchical approach to model observed and unobserved heterogeneity. However, such specification may not separate unobserved heterogeneity from complementary, given that unobserved heterogeneity can also influence the degree to which the marketing actions in one category influence the purchase decision in another category.

Therefore, it is essential to understand the role that complementarity and correlation play in a consumer's observed dependence in purchase behavior across categories. Products of two categories are truly complements if and only if buying them together could bring additional deterministic utility. Gentzkow (2006) developed a discrete model that allows the relationship between pairs of product to be freely estimated from the data and distinguishes between correlation and complementarity.

From a marketer's point of view, it is necessary to understand why products from multiple categories are purchased together by consumers. If a consumer purchases certain products together mainly because of coincidence, promoting in one product category will not bring much additional sales on another category. Understanding if two categories are truly complements can help retailers to manage product assortment and marketing activities. In sum, a purchase timing decision can be modeled through the approach of *whether to buy* if such a decision is evaluated by a utility function, or through the approach of *when to buy* if elapse time or hazard rate is estimated with a parametric (nonparametric) form. In addition, it is necessary to estimate purchase incidence in one category with consideration of purchase incidence in a related category.

Table 3.4

Review of Brand Choice Related Models

Research Interest	Specification	Representative Studies
Purchase Event Feedback	Share of a consumer's purchases of a brand during an earlier holdout period	Tellis (1988) Bucklin and Lattin (1991) Tellis and Zurfryden (1995)
	Exponentially weighted averages of past purchases	Guadagni and Little (1983) Gupta (1988) Tellis (1988) Krishnamurthi and Raj (1988) Chiang (1991) Kannan and Wright (1991) Chintagunta (1993)
	Lagged choice	Bucklin and Lattin (1991) Bucklin and Gupta (1992) Tellis and Zurfryden (1995) Zhang and Krishnamurthi (2004)
	Lagged marketing variable	Zhang and Krishnamurthi (2004) Kumar, Venkatesan, and Reinartz (2006)
Overview of State Dependence	—	Seetharaman (2003) Roy, Chintagunta, and Haldar (1996) Seetharaman (2004)

Brand Choice

Another important decision a consumer has to make is brand choice. Lilien, Kotler, and Moorthy (2003) proposed that most brand choice models differ in how they handle population heterogeneity, purchase-event feedback, and exogenous market factors to be included in the model. The authors defined that difference in marketing factors as reflected in which specific market factors should be included in the model and whether such effects cause structural shift. There are three different types of models in the purchase-event feedback: the zero-order model, the Markov model, and the learning model. The difference among the three types of model depends on to what degree previous purchase history has an effect on brand choice probability. Most models try to capture previous purchase feedback in some sense with various approaches. In this section, we will review how previous research has different approaches of modeling purchase-event feedback and how state dependence is defined and captured (Table 3.4). As shown in Table 3.4, purchase-event feedback can be captured through some brand loyalty measurement based on choice history, lagged brand choices, and lagged marketing variables.

Papatla and Krishnamurthi (1992) suggested that a static brand choice model focuses on brand choice from a set of alternatives in one occasion, whereas a dynamic brand choice model captures repeated choice behavior in multiple purchase occasions. Therefore, when a consumer faces a brand choice decision on each shopping occasion, would he or she form the habit of choosing the same brand? There are logical explanations for such happenings. Consumers may pick the same brand to lower the search cost and reduce risks involved on a purchase occasion. Over time, a consumer may get attached to a brand and repeatedly buy the same brand due to his or her level of brand loyalty. A consumer may also respond more favorably to the marketing of a particular brand than to that of other brands and reinforce his or her positive attitude during each shopping trip.

In order to capture the purchase-event feedback, Guadagni and Little (1983) added the brand loyalty variable to the utility function, assuming that previous choice has a positive effect on future purchase. The brand loyalty variable is measured through the exponential weighted average of past purchases. Therefore, observed heterogeneity in purchase history across households and structural dependence between successive choices are captured to some degree.

Similarly, Gupta (1988) modeled brand choice through a multinomial logit model. The utility of a brand is a function of brand-specific constant, regular brand price, promotional price cut, feature, display, brand loyalty, and size loyalty. Brand loyalty defined in his model is similar to the measurement adopted by Guadagni and Little (1983). Chiang (1991) and Chintagunta (1993) assumed that a brand is chosen if its adjusted price is the lowest among all brands in a product category and is lower than the reservation price. They also captured purchase-event feedback through brand loyalty as defined by Guadagni and Little. Rather than assuming a standard extreme value distribution with the mean as zero and the scale parameter as one, Chiang suggested that the unobserved utilities of the brands follow a generalized extreme value distribution with a parameter controlling for the nonlinearity in relationships among brands, consumer characteristics, and marketing variables. Chintagunta (1993) incorporated unobserved heterogeneity in household intrinsic brand preference through a random coefficient model.

Bucklin and Lattin (1991) suggested that brand choice depends on the purchase state (planned vs. opportunistic). In a planned state, a consumer will choose a brand based on his or her loyalty to the brand, and whether the brand was bought last time in a planned state. Similarly, in an unplanned state, a consumer will choose a brand depending on the price, promotion, and brand loyalty. Therefore, marketing variables only affect brand choice when a consumer is in the opportunistic state. However, in their model, brand loyalty is defined as the share of a consumer's purchase of the brand during an earlier holdout period.

Tellis and Zurfryden (1995) modeled brand choice conditional on purchase incidence and store visits through a multinomial logit model. The utility of a particular brand is a function of brand loyalty, prior purchase, and marketing mix variables in the retailer store. Therefore, they used brand loyalty as well as lagged choice to capture purchase-event feedback. They suggested that cross-sectional heterogeneity can be captured by brand loyalty, whereas lagged choice can model temporal heterogeneity. In addition, the specification of having two terms as compared to a previously used single index of brand loyalty provides the advantage of mathematical tractability (Tellis and Zurfryden 1995).

In order to allow the effect of lagged purchase to change over time for different individuals, Zhang and Krishnamurthi (2004) also examined the time-varying pattern of *variety seeking* and *inertia.* They suggested that variety seeking brings novelty to the consumption experience, whereas inertia saves a consumer time and cognitive effort in decision making. Therefore, they specified inertia as a function of the number of consecutive purchases of the same alternative without promotion up to the previous time period. Similarly, they captured variety seeking as a function of the number of purchases of different alternatives without promotion up to the previous time period. In their model, the response parameter for lagged purchase is a function of *variety seeking* and *inertia.* In addition, they also modeled the effect of lagged promotion on current brand choice probability.

Therefore, how best to capture the effect of the past purchase on current purchase behavior? Seetharaman (2003) suggested that there are two streams of research to capture state dependence: stochastic choice stream and discrete choice stream. The stochastic choice stream uses a probabilistic model to examine dependence, suggesting that the probability of buying a brand depends on either the previous choice or the previous probability of buying the brand,

whereas the discrete choice model approach uses random utility models to estimate temporary dependencies in a consumer's brand choice (Seetharaman 2003). He believes that both streams focus on two sources of state dependence: the lagged choice effect and the lagged evaluation effect. These two effects can be referred to as structural state dependence and habit persistence in econometrics literature (Seetharaman 2003). Therefore, the lagged choice effect is reflected through a consumer's observed repeat purchase behavior, whereas the lagged evaluation effect emphasizes the influence of the past, such as previous marketing activities, on current evaluation of a brand's utility.

Roy, Chintagunta, and Haldar (1996) stated that dynamic choice models should incorporate the influence of marketing stimuli, previous experience, habitual persistence, and inter-individual variations in brand reference and marketing response. Therefore, they suggested that there are three types of temporal dependence: state dependence, habit persistence, and unobserved heterogeneity. They defined structural state dependence or purchase feedback as "the influence of observed past experience (through actual purchase) with a brand, on current choice probabilities." Roy, Chintagunta and Haldar (1996) also defined habit persistence as "the influence of prior propensities to select a brand on current selection probabilities." Usually, structural state dependence is modeled with a lagged choice variable or a function of the lagged choice, whereas habit persistence is measured through the serial correlation in the error components over time (Roy, Chintagunta and Haldar 1996). These authors suggested that unobserved factors such as word-of-mouth communication arrive with a generalized Poisson process. Therefore, the random component of the utility function is not independent as assumed in the popular multinomial logit model, but rather follows a Poisson distribution.

The serial correlation in the random component over time therefore builds the habit persistence that creates dependence in brand evaluation over time. Roy, Chintagunta and Haldar suggested that the stochastic marketing signals experienced by a consumer for a brand at time t follows a multivariate extreme-value distribution, and the underlying process of brand choice is that described by Markov (Roy, Chintagunta, and Haldar 1996). Therefore, the probability of choosing a brand depends on whether the brand was chosen in the previous purchase. One can think of different brands as different states in a Markov chain. The difference between the probability of staying in the same state and that of moving to another state from one purchase occasion to another is the serial correlation between the utility-maximizing alternatives on two occasions, or the serial correlation between the stochastic random shocks of successive time periods.

As already discussed, the serial correlation is used to capture habit persistence, which is formed through the stochastic process of random shocks. Structure dependence refers to the influence of past experience. Therefore, it is partially observed in the data through purchase incidence and usually captured through a lagged purchase indicator. Unobserved heterogeneity is consumer-specific and can be modeled through a parametric or semi-parametric form of consumer-specific intercept or marketing response parameters in the deterministic part of the utility function. According to Roy, Chintagunta, and Haldar (1996), previous experience with a brand may have an overstated impact if habit persistence is not specified. Their results discovered that habit persistence effects diminished greatly after state dependence was accounted for, and state dependence effects dropped significantly once unobserved heterogeneity was estimated.

Seetharaman (2004) further expanded the model by including the lagged effect of marketing variables and decomposing habit persistence. Therefore, he suggested that there are four types of state dependence: structural state dependence, habit persistence types 1 and 2, and carryover effects. He believes structural state dependence can be inertia and variety seeking. Positive structural state dependence is labeled as *inertia*, while negative structural state dependence is referred as

Table 3.5

Review of Purchase Quantity Related Models

Research Interest	Specification	Representative Studies
	No dependence	Gupta (1988)
	Heckman procedure (Maddala approach)	Krishnamurthi and Raj (1988) Tellis (1988)
Dependence between quantity and choice (brand/category)	Single utility function	Chiang (1991) Chintagunta (1993) Arora, Allenby and Ginter (1998)
	Bivariate logit model	Zhang and Krishnamurthi (2004) Niraj, Padmanabhan, and Seetharaman (2007)

variety seeking (Seetharaman 2004). Habit persistence 1 measures the effects of serial correlations between utility-maximizing alternatives on successive purchase occasions of a consumer, which is similar to the habit persistence modeled in Roy, Chintagunta, and Haldar's (1996) model. Habit persistence 2 measures effects of serial correlated error terms in the random utility function. To capture habit persistence 2, Seetharaman added the geometric decay of the previous random error term to the utility function. Such random errors are simulated from an AR(1) process. In addition, he also decomposed the utility function into a function of geometric decay of marketing variables. However, the justification for having two separate habit persistence types is not very clear from the theoretical point of view.

Keane (1997) suggested that distinguishing between state dependence and heterogeneity is very important to a marketer, since state dependence would allow for the effect of promotion to last beyond the promotion period. Seetharaman, Ainslie, and Chintagunta (1999) questioned whether state dependence is a trait of a consumer or a category. They discovered that households have similar state dependence across categories and there are significant correlations between state dependence effects and marketing variables sensitivities. In sum, the key understanding of brand choice behavior is how a consumer's previous experience as well as historic and contemporary market conditions affects the consumer's current choice of a brand. Marketers should fully utilize a consumer's state dependence over time to maximize return on marketing investment. Therefore, it is essential for a marketer to understand what is driving state dependence. For example, if state dependence is mainly determined by past experience, a marketer should use samples to encourage product trial. Similarly, if state dependence is mainly through favorable response to marketing stimuli over time, a marketer should consider advertisement or feature and display. In conclusion, different components of state dependence may mean very different marketing strategies for marketers.

Purchase Quantity

Differences in modeling purchase quantity mostly rely on a researcher's view of the dependence between choice and quantity decisions. In this section, we will review some general approaches of addressing such dependence (Table 3.5). As shown in Table 3.5, the approaches include: correcting for selection bias in the quantity decision through a Heckman procedure or Maddala approach, using a single direct utility to incorporate all three decisions, and proposing a bivariate logit model to jointly estimate purchase quantity and choice decisions.

Gupta (1988) modeled purchase quantity through an ordered regression model. He argued that such a model is more appropriate than the regression model, since the response variable is measured in purchase units. In this model, purchase quantity is a function of average purchase quantity, estimated inventory of the previous purchase, time since previous purchase, and marketing variables such as price, display, and consumer-specific variables. He modeled purchase quantity independent of brand choice decisions.

Tellis (1988) suggested that brand choice and purchase quantity should be modeled jointly through a two-stage process of choice and quantity model, since the values of the dependent variable are censored and only purchased quantity of the preferred brand is observed. Therefore, the random component of brand choice utility may be correlated with the error term of quantity model due to reasons such as common unobserved marketing signals, unobserved heterogeneity, and so forth. Krishnamurthi and Raj (1988) also agree that the decision of how much to buy should be modeled jointly with the decision of which to buy. The modeling of choice and quantity decisions separately could result in biased and inconsistent estimation if the two decisions are not statistically independent (Krishnamurthi and Raj 1988).

Tellis (1988) proposed the Heckman procedure (1979) and Maddala approach (1983) to adjust for the selection bias. He also believed that the two-stage procedure can provide insights in terms of identifying different variables that matter in the choice stage and the quantity stage. For the first stage of the Heckman procedure, a probit model is used to model brand choice. Therefore, the error term in the random component of brand choice follows a normal distribution. However, the random error may not be normally distributed. Krishnamurthi and Raj (1988) presented a model allowing for a general distribution of the error term. The only assumption is that the error term of the quantity function is linear in some specified strictly increasing transformation of the random error in the choice function. The underlying distribution of the error term of the quantity function does not need to be specified unless the maximum likelihood procedure is used to estimate the parameters (Krishnamurthi and Raj 1988).

To ensure that the observed choices provide the highest utility to a consumer, the decisions of purchase quantity, brand choice, and purchase timing should be modeled jointly (Chintagunta 1993). Chintagunta provided a direct utility function with all three decisions incorporated and thereby he assumed that a consumer will make an optimal decision of whether to buy, which to buy, and how much to buy given the price of each brand, quality attributes of each brand, and budget constraint. Since direct utility function is a function of quantity purchased, it can be used to estimate the demand and choice as long as the combined decisions provide the highest utility in the alternative space. Since all three decisions are combined in a single utility function, there is no need to correct for selection bias. Arora, Allenby, and Ginter (1998) also modeled purchase quantity decision jointly with choice decision through a single direct utility function just as Chiang (1991) and Chintagunta (1993) did. They estimated parameters at the individual level through a hierarchical Bayes model rather than a continuous random effects model. Because the observed purchase quantity is discrete in nature, they specified cutoff points for the underlying utility function of purchase quantity. Similar to Chiang's approach, they also used the expenditure share function to find out the reservation value.

According to Zhang and Krishnamurthi (2004), the single utility function approach used by Chiang (1991) and Chintagunta (1993) assumes that the specification of quantity as a function of brand utility imposes substitution between quality and quantity in the choice decision. Therefore, Zhang and Krishnamurthi (2004) proposed a joint probability distribution for the error of the choice/incidence function and that of the quantity function. Since the error term for the choice/incidence function follows a logistic function, they also assumed a logistic function with mean zero and a scale parameter for the quantity random term. The two error terms are therefore assumed to

follow a flexible bivariate logistic distribution by Gumbel (1961). Zhang and Krishnamurthi (2004) also suggested that this specification allows the correlation coefficient of the bivariate logistic distribution to be estimated from the data rather than to be fixed at 0.5. Therefore, the correlation coefficient estimated from the data can tell whether the unobserved factors that affect purchase incidence and those that influence purchase quantity are positive, negative, or unrelated. Using the bivariate logistic distribution also gives them the advantage of having a closed form solution for the likelihood function.

Gönül and Ter Hofstede (2006) modeled order volume and order timing of catalog customers. They believe that order volume depends on the timing of the order and mail and proposed a lognormal distribution for the order volume if there is a purchase, and zero otherwise. They applied Bayesian analysis to capture consumer heterogeneity, but did not correct for selection bias. In conclusion, most researchers believe that the purchase quantity should be modeled jointly with choice decision as long as the purchase quantity decision is determined by the choice decisions of whether to purchase in a category and which brand to choose in the category.

Whom to Target

Once the practitioners have obtained the knowledge of when their customers purchase, which brands/categories they choose, and how much they buy, it is essential for them to optimize their resource allocation based on this knowledge. Therefore, the decision of whom to target is based on the understanding of the customers' heterogeneous behavior in purchase timing, brand choice, and purchase quantity and their differentiated marketing response elasticity on the three decisions. With the advance of information technology and statistical computation, we now have the luxury of obtaining a customer's purchase history, estimating individual-level parameters, and predicting individual purchase behavior and profitability.

Because of limited marketing resources, a firm should strategize how to contact each customer at the individual level and at the aggregate level. For example, retailers can send customized coupons to customers who are more likely to respond to the retailers' promotions. Similarly, manufacturers can contact their customers or send out personalized catalogs at the right time with the right information. Therefore, we believe that it is very important to channel such knowledge that is generated through models we discussed here, from the traditional retailer setting to a broader industry base. In other words, we need not only to help practitioners understand customers' behavior, but also to teach them how to implement resource allocation based on our predictions of customers' future purchase behavior.

Gönül and Ter Hofstede (2006) applied the Bayesian decision rule to design an optimal mailing schedule for each individual customer in a catalog company. The key is to maximize the expected utility defined as below:

$$E(U(\Theta_i, \hat{D}_i) \mid T_i, Z_i, D_i) = \int_{\Theta_i} U(\Theta_i, \hat{D}_i) p(\Theta \mid T_i, Z_i, D_i) d\Theta_i \tag{1}$$

where $U(\Theta_i \hat{D}_i)$ is the utility function depended on the customer level parameter set Θ and a mailing schedule D, and $P(\Theta|T_i, Z_i, D_i)$ is the posterior distribution of the parameters.

Kumar, Venkatesan, and Reinartz (2006) proposed that the likelihood that a customer i will purchase product j at time t is defined as:

$$L_i = \prod_{r_i=1}^{R_i} [\{ \prod_{j=1}^{J} (f_i(t,j)^{\delta_{ijt}} \}^{C_{r_i}} S_i(t)^{1-C_{r_i}}] \quad (2)$$

where $f_i(t,j)$ is customer i's probability density function of purchasing product j, and $S_i(t)$ is the survival function of purchase probability for customer i. They also suggested using Bayesian estimation to eliminate sampling error. Once firms obtain a better prediction of a customer's purchase pattern, they can contact only those customers who are more likely to buy in that period with the right message. In a nutshell, firms can decide whom to target, when to sell, and what to sell based on the prediction.

Fundamental Differences in Various Approaches

Consumer's Decision Process

After reviewing various models, we noticed that the difference in modeling approaches reflects the researchers' different perceptions of the reality that they were trying to model. The first difference is their assumption of a consumer's decision process. Gupta (1988) proposed that purchase incidence and brand choice can be modeled separately because brand choice is determined by the marketing condition of the week when a consumer is going to make a purchase.

Another assumption is that a consumer goes through a sequential decision-making process. Therefore, the three decisions can be modeled through conditional probability. Under such an assumption, a consumer will decide whether or not to make a category purchase. This decision usually depends on the consumer's intrinsic demand for the product, which is usually controlled through the consumption rate and inventory level. Once a consumer decides to make a purchase in a product category, he or she will choose the brand with the highest utility. This utility is determined by marketing variables of various brands in the category and consumer specific characters such as brand preference, loyalty, and so forth. The final decision a consumer has to make is purchase quantity. Usually, a consumer's decision of how much to buy is determined by his or her inventory level. Brand choice and brand incidence can be modeled through a nested logit model. Similarly, purchase quantity can be modeled conditional on brand choice with correction for selection bias.

Another assumption about a consumer's decision process is that consumers make simultaneous decisions of *whether to buy*, *which to buy*, and *how much to buy* subject to the budget constraint so that the utility will be maximized in each shopping occasion. Of course, one can think of purchase quantity as the secondary demand derived from the primary demand of brand choice. This specification is important to examine the role played by marketing variables. In this situation, marketing variables can simultaneously influence the three decisions. One approach (Arora, Allenby, and Ginter 1998; Chiang 1991; Chintagunta 1993) is to capture them through a direct utility function, which also incorporates purchase quantity and reservation price. Reservation price allows a consumer to compare the option of buying versus not buying in the category. Purchase quantity ensures that a consumer's decision of *how much to buy* maximizes the overall utility on the shopping occasion. However, unlike models assuming a sequential decision process, price appears as an element of budget constraint.

Another approach is to specify a bivariate distribution for the error component of the choice function and that of the quantity function. If the decisions of *whether to buy*, *which to buy*, and *how much to buy* are subject to the commonly unobserved factors including unobserved marketing signals, such an interrelationship allows us to model them together through a joint distribution.

Unobserved Heterogeneity

Another major difference in the various modeling approaches is to what degree unobserved heterogeneity should be captured and how it should be modeled. The simple multinomial logit model assumes that observations within subjects are independent, since the random terms are distributed as IID extreme value distribution. However, such an assumption is hardly realistic. If heterogeneity is not incorporated, observations of multiple purchases from the same household will be regarded as purchases from different households.

Some earlier models addressed observed heterogeneity through variables measured from previous purchase behavior or demographics. However, Chintagunta, Jain, and Vilcassim (1991) suggested that such a practice may affect the estimation of marketing variables if the effect of previous marketing activities is not adjusted from the previous purchase behavior. Another issue is that such a variable might also contain information such as previous random demand shock, which may be stochastic over time. Therefore, including previous purchase behavior as a proximate measure of brand loyalty without adjusting the stochastic nature of random demand shocks could cause endogeneity. Finally, observed heterogeneity has only limited ability to capture individual difference.

Chintagunta (1993) suggested that the estimation for the coefficients of marketing variables could be biased if unobserved heterogeneity is not modeled. Perhaps one of the most important reasons to model unobserved heterogeneity is to understand consumer behavior at the individual level. Allenby and Rossi (1999) suggested that marketers need to understand the diversity of preferences and sensitivities that exist in the market to offer differentiated products and marketing communications. Therefore, previous literature has addressed heterogeneity in various ways.

Kamakura and Russell (1989) captured heterogeneity through a number of latent segments. Therefore, they assume that consumers within a segment are homogenous in terms of marketing response elasticity and consumers from different segments are heterogeneous. Each consumer has various probabilities of belonging to different segments, and such probabilities are estimated from the data so that the overall likelihood is maximized. The number of segments in the market is determined by the model-fitting criterion. Although this approach is effective in segmenting the market, it has limitations in understanding individual behavior.

Chintagunta, Jain, and Vilcassim (1991) discussed the importance of modeling unobserved heterogeneity and suggested an alternative approach to capture heterogeneity in brand choice behavior by assuming that a consumer's preference for each brand is stable over time and can be captured through a random or fixed coefficient model. If a parametric distribution is not appropriate, a finite mixture model with a number of support points estimated from the data can be used (Chintagunta, Jain, and Vilcassim 1991). However, Arora, Allenby and Ginter (1998) suggested that the finite mixture model cannot capture tail behavior very well.

In addition to applying a parametric or semi-parametric random coefficient model-to-model heterogeneity across consumers, one can also use Bayesian statistics by finding the posterior distribution of consumer-specific parameters. For example, Gönül and Ter Hofstede (2006) modeled the individual-level hazard function by specifying a measure of customer-specific parameter through Bayesian estimation.

Allenby and Rossi (1999) suggested that the random coefficients model can only provide an estimation of hyper-parameters of underlying population distribution that captures heterogeneity, whereas the Bayesian model could allow for individual-level estimation and capture uncertainty existing in individual level parameters. However, a specification other than normal

distribution for the random error could lead to a complicated posterior distribution with much difficulty to sample from.

Andrews, Ainslie, and Currim (2002) compared the performances of finite mixture models and hierarchical Bayesian models subject to changes in (1) the number of mixture components, (2) distance between mixture components, (3) underlying distribution of parameters within mixture components, (4) variance of parameters within mixture components, (5) the number of households and purchase per household, and (6) error variance. They discovered that despite the finding that HB-estimated models fit better than FM models, both models are equally effective in parameter estimation and prediction even with the violation of the underlying assumptions (Andrews, Ainslie, and Currim 2002). However, they also warned that when the observed number of purchases is too small, the HB approach cannot identify individual parameters very well. Bayesian approach generally has a shrinkage effect allowing for the estimation to be more reliant on aggregate household information or prior information than limited individual observation. However, Andrews, Ainslie, and Currim (2002) observed poor parameter recovery and predicative accuracy with a limited number of individual purchase observations.

In terms of estimating various models, maximum likelihood procedure is a common approach if Bayesian analysis is not applied. The nested logit model is usually sequentially estimated. Bucklin and Gupta (1992) listed the advantage and disadvantage of sequential estimation: the advantage is the computation tractability; the disadvantage is being less efficient than simultaneous estimation. They adopted sequential estimation in their study by maximizing the likelihood according to brand choice and then purchase incidence. In summary, the assumption of a consumer's underlying decision-making process determines the fundamental modeling approach used by a researcher. In addition, whether consumer heterogeneity should be captured and how it should be captured reflect a researcher's interest of modeling individual behavior.

Modeling at Least Two of the Decisions

Therefore, if a researcher is interested in modeling two of the three decisions, what are the general approaches that can be used? Table 3.6 provides some general thoughts. As shown in the table, there are various combinations of modeling approaches for estimating at least two decisions. For example, timing and brand choice can be modeled through a nested logit, hierarchical latent regression, and so on. Purchase quantity can be estimated using a regression with selection bias corrected, along with purchase timing modeled with a logit model or a probit model. If purchase quantity is incorporated in the utility function, all three decisions can be captured through a single direct utility function.

To jointly model the purchase timing/incidence decision and brand choice decision, one can model the latent utility of making a purchase in the product category and choosing a particular brand through a nested logit model or similar specification. Usually, the dependence between purchase incidence and brand choice is nested through a category-specific value. Bucklin and Gupta (1992) stated that category value captures the attractiveness of the product category due to price and promotion activities on the individual brands in the category. They suggested that the probability of household h making a category purchase at time t is:

$$P_t^h(inc) = \frac{\exp(V_t^h)}{1+\exp(V_t^h)} \tag{3}$$

Table 3.6

Model Specifications

Decisions Studied	Model Specification(s)	Representative Studies
	Nested logit	Bucklin and Lattin (1991) Bucklin and Gupta (1992)
Timing/incidence and brand choice	Hierarchical latent regression	Ainslie and Rossi (1998)
	Hazard (timing) Probit (choice)	Kumar, Venkatesan, and Reinartz (2006)
Brand choice and purchase quantity	Probit/logit (choice)	Krishnamurthi and Raj (1988)
	Regression with selection bias corrected (quantity)	Tellis (1988)
	Regression (elapse time) Regression (quantity)	Neslin, Henderson, and Quelch (1985)
	CM-Poisson/logit (timing) Log normal (quantity)	Boatwright, Borle, and Kadane (2003)
Timing/incidence and purchase quantity	Hazard function (timing) Log normal (quantity)	Gönül and Ter Hofstede (2006)
	Bivariate logit (timing) Conditional bivariate logit (quantity)	Niraj, Padmanabhan, and Seetharaman (2007)
	Erlang (timing) Logit (choice)	Gupta (1988)
Timing/incidence and purchase quantity and brand choice	Ordered logit (quantity) Single direct utility function subject to budget constraint	Chiang (1991) Chintagunta (1993) Arora, Allenby, and Ginter (1998) Song and Chintagunta (2006)
	Bivariate logit	Zhang and Krishnamurthi (2004)

where V_t^h is the deterministic component of category purchase utility to household *h* at time *t*. The conditional probability of household *h* choosing brand *i* at time *t* is:

$$P_t^h(i|inc) = \frac{exp(U_{it}^h)}{\sum_k exp(U_{kt}^h)} \tag{4}$$

where U_{it}^h is the deterministic component of brand *i*'s utility to household *h* at time *t*.

Ainslie and Rossi (1998) suggested that brand choice and purchase incidence can be modeled jointly through a hierarchical latent regression. They specified the model as:

$$I_{h,c,t} = r_c, \ \ where \ y_{h,c,t,r_c} = \max_j (y_{h,c,t,j})$$
$$j = 1,..., n_c; c = 1,..., C; h = 1,..., H, \tag{5}$$

where n_c is the number of brands in category *c*, $I_{h,c,f}$ is the brand chosen by household *h* in category *c* at time *t*, r_c is the *r*th brand in category *c*, and y_{h,c,t,r_c} is the latent utility of household *h* purchasing the *r*th brand in category *c* at time *t*.

The decisions of brand choice and purchase quantity can be modeled jointly through a bivariate logit function if the purchase quantity is assumed to be discrete. If the purchase quantity is continuous, it can be estimated using regression. However, it is necessary to correct for the selection bias with Heckman procedure. Just as Krishnamurthi and Raj (1988) suggested, the continuous dependent variable is limited in range and depends on the selection of one of the alternatives. In their model, they first specified that the latent utility for a binary choice alternative is:

$$I_i^* = Z_i\gamma - \varepsilon_i \tag{6}$$

where γ is a vector of choice parameters. The observed purchase quantity is:

$$Y_{1i} = X_{1i}\beta_1 + u_{1i} \text{ if } \varepsilon_i \le Z_i\gamma, \text{ i.e., } I_i = 1 \tag{7}$$

Since the choice model error ε_i may be correlated with the error of the quantity model u_{1i}, the regression model needs to be adjusted as:

$$Y_{1i} = X_{1i}\beta_1 - \alpha_1 S_{1i} + v_{1i} \tag{8}$$

where S_{1i} is the selection bias, which equals to $\phi(Z_i\gamma)/\Phi(Z_i\gamma)$, and v_{1i} is the error. For more details regarding how to adjust for bias if the choice is not modeled through a probit function, please refer to Krishnamurthi and Raj (1988).

To jointly model purchase incidence and purchase quantity, bivariate logit is one of the options if the model specification for both quantity and incidence is logit. Gönül and Ter Hofstede (2006) specified that purchase quantity follows a log normal distribution conditional on purchase incidence decision. Since the parameters from both the log normal distribution and hazard function are customer specific and follow a normal distribution, unobserved customer-specific characteristics that influence both purchase incidence and quantity decisions can be captured. They specified that the volume of an order placed by customer i at time t of order spell k, $Z_{ik}(t)$, has mean $u_{ik}(t)$ and variance σ_z^2. Therefore, the conditional distribution of $Z_{ik}(t)$is defined as:

$$[Z_{ik}(t)\big|T_{ik},\ \mu_{ik}(t),\ \sigma_z^2] \sim \begin{cases} LogN(u_{ik}(t),\ \sigma_z^2) & \text{if } T_{ik} = t \\ I(Z_{ik}(t) = 0) & \text{otherwise} \end{cases} \tag{9}$$

Boatwright, Borle, and Kadane (2003) also suggested that purchase quantity follows a log normal model with the mean value depending on inter-purchase time.

If we need to model three decisions together, a single direct utility function can be used to capture the three decisions so that the utility can be maximized within a consumer's budget constraint. Chintagunta (1993) suggested that the utility maximization problem of household i at store visit t can be expressed as

$$\max_{Q_{ijt}} \ u\Big(\sum_{j=1}^{N} \psi_{ijt}(X_{ijt}, \varepsilon_{ijt}) Q_{ijt}, Z_{it}\Big) \quad (\textit{objective function}) \text{ subject to} \tag{10}$$

$$\sum_{j=1}^{N} p_{ijt} Q_{ijt} + z_{it} = y_{it} \quad (\textit{expenditure equation}) \tag{11}$$

$$Q_{ijt} \geq 0, \; Z_{it} > 0 \quad (\textit{nonnegativity constraints}) \tag{12}$$

where p_{ijt} is price for brand j faced by household i, and y_{it} is the total expenditure of household i on store visit t. He defined ψ_{ijt}, the quality attribute of brand j to household i at time t, as:

$$\psi_{ijt} = \exp(\gamma_{ij} + \sum_{s} X_{ijts}\beta_s + \varepsilon_{ijt}), \quad s = 1,2,...,S, \tag{13}$$

where γ_{it} is household i's time-invariant intrinsic preference toward brand j. Therefore, in this model specification, purchase quantity is also part of the utility function.

Once again, bivariate logit can be used to model the three decisions. Zhang and Krishnamurthi (2004) derived the probability of household i purchasing the quantity of alternative k at time t , q_{ikt} as:

$$Pr(I_{it} = 1, \; B_{ikt} = 1, \; Q_{ikt} = q_{ikt}) = \frac{e^{V_{ikt}}}{\sum_{j=0}^{k} e^{V_{ijt}}} \frac{\delta_\xi e^{\delta_\xi (Z_{ikt}\phi_i - q_{ikt})}}{[1 + e^{\delta_\xi (Z_{ikt}\phi_i - q_{ikt})}]^2}$$
$$.[1 + \theta(1 - \frac{e^{V_{ikt}}}{\sum_{j=0}^{k} e^{V_{ijt}}}) \frac{-1 + e^{\delta_\xi (Z_{ikt}\phi_i - q_{ikt})}}{1 + e^{\delta_\xi (Z_{ikt}\phi_i - q_{ikt})}}] \tag{14}$$

where V_{ikt} is the systematic utility of alterative k to household i at time t; δ_ξ is the scale parameter of $\varepsilon_{ikt} = \max_{j=0,1,..K \text{ and } j \neq k} \{V_{ijt} + \varepsilon_{ijt}\} - \varepsilon_{ikt}$, which follows an extreme value distribution; Z_{ikt} is a vector of explanatory variables for purchase quantity; θ is the correlation between ε^*_{ikt} and ξ_{ikt}, the random utility of alterative k to household i at time t.

Gupta (1988) modeled the decisions separately by estimating purchase quantity through an ordered logit, brand choice through a logit, and timing decision through an Erlang distribution. In his model, he suggested that the probability density function of inter-purchase time t for consumer i in week w is defined as:

$$f_{iw}(t) = \alpha_{iw}^2 t \exp(-\alpha_{iw} t) \tag{15}$$

$$\alpha_{iw} = \exp(-c' Y_{iw}) \tag{16}$$

where Y_{iw} is a vector of explanatory variables that may determine the inter-purchase time. In summary, various approaches can be chosen depending on (1) how the purchase timing decision should be captured, (2) how brand choice function is specified, (3) how purchase quantity is estimated and whether it is discrete or continuous, and (4) how to address the interdependence of the three decisions.

Applications of Purchase Timing, Quantity, and Choice Decision Models

Most of the research on quantity, choice, and timing models has been conducted in the business-to-consumer context. There are several streams of research with issues addressed as shown in Table 3.7. We also listed representative studies of each stream in Table 3.7.

The first stream of research is to understand the effect of marketing variables on the three decisions in the form of purchase acceleration, stockpiling, and brand switching. Manufacturers are mostly interested in how marketing variables influence brand switching, whereas retailers are more concerned about category purchase incidence. Neslin, Henderson, Quelch (1985) examined the differentiated effects of coupons, manufacturer advertising, retailer advertising, and price change on consumers' inter-purchase time and purchase quantity counting for consumer heterogeneity in loyalty and other characteristics. They found out that increased purchase quantity rather than shortened inter-purchase time is more likely to occur in terms of purchase acceleration.

Gupta (1988) modeled brand choice, purchase timing, and quantity to understand the effects of sales promotion and decomposed the *sales bump* based on these three decisions. The results suggested that an increase in sales due to promotion mostly comes from brand switching rather than purchase timing acceleration and stockpiling. Krishnamurthi and Raj (1988) modeled the effects of price on purchase quantity and brand choice decisions jointly and found out that competitive pricing has a greater effect on brand choice than on purchase quantity. Tellis (1988) tested the effects of advertising exposure on brand choice subjecting to the mediating role brand loyalty played. He found that feature, display, and price have a stronger effect on choice decision than advertising does.

The second stream of research focused on revealing the structure of consumer heterogeneous response toward marketing variables. Kannan and Wright (1991) found the existence of a structured market and proposed a method to reveal such structured accounting for consumers' heterogeneous responses toward marketing variables. Similarly, Bucklin and Gupta (1992) segmented consumers based on brand choice as well as category purchase incidence, since some customers respond to marketing variables through brand switching, some through purchase acceleration, some through both, and the rest do not respond at all.

The third stream of research further integrated the three decisions through a single direct utility function of purchase timing, brand choice, and quantity, whereas most of the traditional approaches modeled the three decisions through the joint probability function. Chintagunta (1993) suggested that modeling the three decisions separately cannot ensure the maximal utility by a consumer and therefore a single direct utility function should be used. Furthermore, the utility function he proposed can examine the impact of marketing variables unconditional on purchase incidence with unobserved heterogeneity in reservation prices and brand preferences (Chintagunta 1993). He found out that the marketing variables have differentiated effects on the three decisions. Chiang (1991) examined the decisions of *whether to buy*, *what to buy*, and *how much to buy* with a single direct utility function considering the nonpurchase probability. He found out that promotion and past purchase history positively affect brand choice and purchase quantity.

The fourth stream of research focused on multicategory purchase dependence. Ainslie and Rossi (1998) measured correlations in marketing sensitivities across multiple categories through observed heterogeneity such as income, family size, and shopping behavior as well as unobserved heterogeneity. Chintagunta and Haldar (1998) also examined the dependence of purchase incidence in two related product categories (complementarity or substitutability) through a bivariate hazard function. Similarly, Manchanda, Ansari, and Gupta (1999) proposed

Table 3.7

Research Interests

Research Interest(s)	Representative Studies
Decomposing the effect of marketing variables on the choice and quantity decisions	Gupta (1988) Neslin, Henderson and Quelch (1985) Krishnamurthi and Raj (1988) Tellis (1988)
Market segmentation based on response to marketing variables through the choice and/or quantity decisions	Kannan and Wright (1991) Bucklin and Gupta (1992)
Simultaneous estimation of whether to buy, which to buy, and how much to buy	Chiang (1991) Chintagunta (1993) Arora, Allenby, and Ginter (1998)
Cross-category dependence	Manchanda, Ansari, and Gupta (1999) Song and Chintagunta (2006) Ainslie and Rossi (1998) Chib, Seetharaman and Strijnev (2002)
State dependence	Roy, Chintagunta, and Haldar (1996) Seetharaman (2004)
Marketing resource optimization based on the prediction of the choice and quantity decisions	Zhang and Krishnamurthi (2004) Gönül and Ter Hofstede (2006) Kumar, Venkatesan, and Reinartz (2006)

that there are three factors that could influence multicategory dependence: complementarity, co-incidence, and heterogeneity, where the complementary effect is measured through the deterministic part of the utility functions, co-incidence is captured through the covariance matrix of the random part of the function, and unobserved heterogeneity is modeled through random coefficients.

Song and Chintagunta (2006) used aggregate store-level scanner data to examine cross-category effects of marketing variables with brand-specific utility modeled as a function of intrinsic brand preference, price and deal variables, and store and time demand dummy variables, and with a category-specific component of bundled utility modeled as a function of the marginal effect of including a category and the synergistic effect of two categories.

The fifth stream of research examined a consumer's brand choice behavior over time through state dependence and variety-seeking behavior. Seetharaman (2004) modeled four types of state dependence in brand choice: (1) effects of lagged choice, (2) marketing variables carryover effects, (3) lagged random error, and (4) serial correlation in the consumer's utility-maximization alternatives on successive purchase occasions. Therefore, the probability of a consumer choosing a brand is determined not only by a utility function with previous purchase behavior and marketing variables, but also by a switching (or habit persistence) probability.

The last stream of research is mostly concerned with how to apply the brand choice, purchase timing, and quantity model to optimize a manufacturer or a retailer's profitability. Tellis and Zufryden (1995) tried to optimize timing and depth of retail discount with an integrated consumer response model of three decisions, and a retailer model with consumer purchase and retailer inventory. Gönül and Ter Hofstede (2006) derived a Bayesian optimal mailing rule for a catalog company based on a response model of order timing and volume. Kumar, Venkatesan, and Reinartz (2006) demonstrated how to use predictions of purchase timing and product choice to strategize what to sell, when, and to whom.

Future Challenges

Despite the tremendous progress made in the past several decades, modeling these three decisions still provides plenty of opportunities and inspirations for researchers. Table 3.8 lists the future challenges that could push this study further and provide more insightful intelligence for marketers in decision making.

As we see from Table 3.8, an issue worth noticing is endogeneity in choice models. The presence of common marketing signals that are unobserved to researchers may create endogeneity in a simple multinomial logit model, which assumes that the marketing variables are independent of the random component of utility function. Villas-Boas and Winer (1999) suggested that researchers justify the act of not accounting for endogeneity in brand choice models by stating that the marketing-mix variables are common across consumers. However, they suggested that such an assumption of independence in random utility across individuals may be invalid given the unobserved common shocks and word-of-mouth or fashion (wearout) effects.

Villas-Boas and Winer (1999) also suggested that another source of endogeneity is from some marketing variables such as price. They used lagged price as an instrument for price. However, they suggested that the price model error may be correlated to the consumer common error, since the price model error includes unobserved shocks on costs and demand (Villas-Boas and Winer 1999). The results from their study demonstrate that price is indeed endogenous.

Most of the models we reviewed are stationary in parameters. More research should be encouraged to examine the long-term effect of marketing variables that shift the demand structure. Sriram, Chintagunta and Neelamegham (2006) introduced time-varying parameters to the brand choice model and therefore captured the effect of changes in the portfolio of various models and the dynamics in intrinsic brand preference on choice behavior. Overall, there is more to discover in terms of choice decision.

Traditionally, the models discussed are examined in the retailer setting. Kumar, Venkatesan, and Reinartz (2006) showed how to use data from a company's customer relationship management system to track which customer is more likely to buy, when, and what. The knowledge generated from this stream of literature can be channeled into a broader industry base.

The data used in most of the studies are panel data. However, the transaction data owned by firms contain only partial purchase history. Therefore, data augmentation methods such as Bayesian or EM algorithm need to be used to infer the missing purchase information. One of the major challenges is thus to obtain a better estimate of parameters, given only partial information.

Most of the studies are conducted in the business to consumer (B2C) context. However, it is essential to examine choice behavior in the business to business (B2B) context. Since high-tech B2B purchase differs from regular household purchase in that buyers bear more risk in purchasing new products and the switching cost is higher, the behavior assumption in the B2C context may not hold in the B2B context. State dependence and cross-category dependence may also be very different in this setting.

In conclusion, the purchase timing, choice, and quantity decisions have generated much interest in the past several decades. Extensive research has been done to recover the inter-relationships of the three decisions over time and across categories. The increasing complexity in modeling approach allows us to capture more effects and have a deeper understanding of the phenomena. This study provides a preliminary review of various studies and some thoughts regarding the researchers' perception of the problem. Future research can hopefully use this study to determine which set of models are more appropriate for what decisions, and work toward expanding the scope in both the modeling and substantive contribution.

Table 3.8

Future Challenges

Future Challenges	Representative Studies Addressing the Challenges
Address endogeneity issue in modeling purchase timing, choice, and quantity	Villas-Boas and Winer (1999)
Introduce nonstationary parameters to the model	Sriram, Chintagunta, and Neelamegham (2006)
Expand the modeling of three decisions from the retailing setting to a broader industry base	Gönül and Ter Hofstede (2006) Kumar, Venkatesan, and Reinartz (2006)
Model the three decisions with transactional data rather than panel data	Boatwright, Borle, and Kadane (2003) Zhang and Krishnamurthi (2004) Gönül and Ter Hofstede (2006) Kumar, Venkatesan, and Reinartz (2006)
Capture a customer's purchase behavior in the business-to-business context	Kumar, Venkatesan, and Reinartz (2006)

Acknowledgments

The authors thank Joseph Pancras, S. Sriram, and Rajkumar Venkatesan for providing valuable comments on this paper.

References

Ainslie, Andrew, and Peter E. Rossi. 1998. "Similarities in Choice Behavior Across Product Categories." *Marketing Science* 17 (2): 91–106.

Allenby, Greg M., and Peter E. Rossi. 1999. "Marketing Models of Consumer Heterogeneity." *Journal of Econometrics* 89: 57–78.

Andrews, Rick L., Andrew Ainslie, and Imran S. Currim. 2002. "An Empirical Comparison of Logit Choice Models with Discrete Versus Continuous Representations of Heterogeneity." *Journal of Marketing Research* 39 (4): 479–487.

Arora, Neeraj, Greg M. Allenby, and James L. Ginter. 1998. "A Hierarchical Bayes Model of Primary and Secondary Demand." *Marketing Science* 17: 29–44.

Bijwaard, Govert E. 2006. "Regularity in Individual Shopping Trips: Implications for Duration Models in Marketing." Working paper, Econometric Institute, Erasmus University, Rotterdam.

Boatwright, Peter, Sharad Borle, and Joseph B. Kadane. 2003. "A Model of the Joint Distribution of Purchase Quantity and Timing." *Journal of the American Statistical Association* 98 (463): 564–572.

Bucklin, Randolph E., and Sunil Gupta. 1992. "Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach." *Journal of Marketing Research* 29 (2): 201–215.

Bucklin, Randolph E., and James M. Lattin. 1991. "A Two-State Model of Purchase Incidence and Brand Choice." *Marketing Science* 10 (1): 24–39.

Chatfield, C., and G.J. Goodhardt. 1973. "A Consumer Purchasing Model with Erlang Inter-Purchase Times." *Journal of the American Statistical Association* 68 (344): 828–835.

Chiang, Jeongwen. 1991. "A Simultaneous Approach to the Whether, What and How Much to Buy Questions." *Marketing Science* 10 (4): 297–315.

Chib, Siddhartha, P.B. Seetharaman, and Andrei Strijnev. 2002. "Analysis of Multi-Category Purchase Incidence Decisions Using IRI Market Basket Data." *Advances in Econometrics* 16: 57–92.

Chintagunta, Pradeep K. 1993. "Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households." *Marketing Science* 12 (2): 184–208.

Chintagunta, Pradeep K., and Sudeep Haldar. 1998. "Investigating Purchase Timing Behavior in Two Related Product Categories." *Journal of Marketing Research* 35 (1): 43–53.

Chintagunta, Pradeep K., Dipak C. Jain, and Naufel J. Vilcassim. 1991. "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data." *Journal of Marketing Research* 33 (November): 417–428.

Ehrenberg, A.S.C. 1959. "The Pattern of Consumer Purchases." *Applied Statistics* 8: 26–41.

Gentzkow, Matthew. 2006. "Valuing New Goods in a Model with Complementarity: Online Newspapers." Working paper, Graduate School of Business, University of Chicago.

Gönül, F. Füsun, and Kannan Srinivasan. 1993. "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues." *Marketing Science* 12 (3): 213–229.

Gönül, F. Füsun, and Frenkel Ter Hofstede. 2006. "How to Compute Optimal Catalog Mailing Decisions." *Marketing Science* 25 (1): 65–74.

Guadagni, Peter, M., and John D.C. Little. 1983. "A Logit Model of Brand Choice Calibrated on Scanner Data." *Marketing Science* 2 (3): 203–238.

Gumbel, Emil J. 1961. "Bivariate Logistic Distribution." *Journal of the American Statistical Association* 56: 335–349.

Gupta, Sunil. 1988. "Impact of Sales Promotions on When, What, and How Much to Buy." *Journal of Marketing Research* 25 (4): 342–355.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47: 153–61.

Jain, Dipak C., and Naufel J. Vilcassim. 1991. "Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach." *Marketing Science* 10: 1–23.

Kamakura, Wagner, and Gary J. Russell. 1989. "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure." *Journal of Marketing Research* 26: 379–390.

Kannan, P.K., and Gordon P. Wright. 1991. "Modeling and Testing Structured Markets: A Nested Logit Model." *Marketing Science* 10 (1): 58–82.

Keane, Michael P. 1997. "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior." *Journal of Business & Economics Statistics* 15 (3): 310–327.

Krishnamurthi, Lakshman, and S.P. Raj. 1988. "A Model of Brand Choice and Purchase Quantity Price Sensitivity." *Marketing Science* 7 (1): 1–20.

Kumar, V., Rajkumar Venkatesan, Tim Bohling, and Denise Beckmann. 2006. "The Power of CLV: Managing Customer Value at IBM." Working paper, Department of Marketing, University of Connecticut.

Kumar, V., Rajkumar Venkatesan, and Werner Reinartz. 2006. "Knowing What to Sell, When, and to Whom." *Harvard Business Review* (March): 1–8.

Lilien, L. Gary, Philip Kotler, and K. Sridhar Moorthy. 2003. *Marketing Models.* New Delhi: Prentice Hall.

Ma, Yu, and P.B. Seetharaman. 2006. "The Multivariate Logit Model for Multicategory Purchase Incidence Outcome." Working paper, Department of Marketing, Rice University.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press.

Manchanda, Puneet, Asim Ansari, and Sunil Gupta. 1999. "The 'Shopping Basket': A Model of Multicategory Purchase Incidence Decisions." *Marketing Science* 18 (2): 95–114.

Neslin, Scott A., Caroline Henderson, and John Quelch. 1985. "Consumer Promotions and the Acceleration of Product Purchases." *Marketing Science* 4 (2): 147–165.

Niraj, Rakesh, V. Padmanabhan, and P.B. Seetharaman. 2007. "A Cross-Category Model of Households' Incidence and Quantity Decisions." *Marketing Science* (forthcoming).

Papatla, Purushottam, and Lakshman Krishnamurthi. 1992. "A Probit Model of Choice Dynamics." *Marketing Science* 11 (2): 189–205.

Roy, Rishin, Pradeep K. Chintagunta, and Sudeep Haldar. 1996. "A Framework for Investigating Habits, 'The Hand of the Past,' and Heterogeneity in Dynamic Brand Choice." *Marketing Science* 15 (3): 280–299.

Seetharaman, P.B. 2003. "Probabilistic Versus Random-Utility Models of State Dependence: An Empirical Comparison." *International Journal of Research in Marketing* 20: 87–96.

———. 2004. "Modeling Multiple Sources of State Dependence in Random Utility Models: A Distributed Lag Approach." *Marketing Science* 23 (2): 263–271.

Seetharaman, P.B., Andrew Ainslie, and Pradeep Chintagunta. 1999. "Investigating Household State Dependence Effects Across Categories." *Journal of Marketing Research* 36 (4): 488–500.

Seetharaman, P.B., Siddhartha Chib, Andrew Ainslie, Peter Boatwright, Tat Chan, Sachin Gupta, Nitin Mehta, Vithala Rao, and Andrei Strijnev. 2005. "Models of Multi-Category Choice Behavior." *Marketing Letters* 16: 239–254.

Seetharaman, P.B., and Pradeep K. Chintagunta. 2003. "The Proportional Hazard Model for Purchase Timing: A Comparison of Alternative Specification." *Journal of Business & Economic Statistics* 21 (3): 368–382.

Song, Inseong, and Pradeep K. Chintagunta. 2006. "Measuring Cross-Category Price Effects with Aggregate Store Data." *Management Science* 52 (10): 1594–1609.

Sriram, S., Pradeep K. Chintagunta, and Ramya Neelamegham. 2006. "Effects of Brand Preference, Product Attributes, and Marketing Mix Variables in Technology Product Markets." *Marketing Science* 25 (5): 440–456.

Tellis, J. Gerard. 1988. "Advertising Exposure, Loyalty, and Brand Purchase: A Two-Stage Model of Brand Choice." *Journal of Marketing Research* 25 (2): 134–144.

Tellis, J. Gerard, and Fred S. Zurfryden. 1995. "Tackling the Retailer Decision Maze: Which Brands to Discount, How Much, When and Why?" *Marketing Science* 14 (3): 271–299.

Villas-Boas, J. Miguel, and Russell S. Winer. 1999. "Endogeneity in Brand Choice Models." *Management Science* 45 (10): 1324–1338.

Wheat, Rita, D., and Donald G. Morrison. 1990. "Assessing Purchase Timing Models: Whether or Not Is Preferable to When." *Marketing Science* 9 (2): 162–70.

Zhang, Jie, and Lakshman Krishnamurthi. 2004. "Customizing Promotions in Online Stores." *Marketing Science* 23 (4): 561–578.

CHAPTER 4

BRAND EXTENSION RESEARCH

A Cross-Cultural Perspective

MICHAEL A. MERZ, DANA L. ALDEN, WAYNE D. HOYER, AND KALPESH KAUSHIK DESAI

Abstract

It is common in both national and foreign markets that companies leverage strong brand equity by extending their brands into other (congruent and/or incongruent) product categories. While the literature on national brand extensions has developed extensively since the late 1980s and early 1990s, the literature involving brand extension strategies in foreign markets remains very limited. This limits our understanding of effective brand extension strategy across cultures. To facilitate cross-cultural brand extension research, we propose a new conceptual framework regarding effective brand extension strategy in a cross-cultural context. Specifically, we propose ways in which culture might affect consumers' brand extension evaluations. In so doing, we develop several propositions regarding cross-cultural brand extension strategy as a potential agenda for future research. Managerial implications are also discussed.

Introduction

Business Week and the brand consultancy Interbrand annually publish rankings of the top 100 global brands (Interbrand 2006). To be included, a brand must have—among other criteria—a minimum brand value of US$2.7 billion and achieve one-third or more of its earnings outside of its home country. Among the most successful global brands are such icons as Coca-Cola, Microsoft, IBM, Toyota, McDonald's, BMW, Samsung, Google, Dell, Apple, and Ikea. Many of these top global brands have estimated values far in excess of the minimum required for membership in this elite group. For example, Coca-Cola is valued at US$67 billion, Microsoft at US$57 billion, and Samsung (ranked 20th) at US$16 billion (*Business Week* 2006).

Such high levels of brand equity help explain why academics and practitioners have become increasingly interested in global branding. The development of new technology, labor mobility, cross-border tourism, and the abolition of trade barriers further increase the strategic attractiveness of global branding (Holt, Quelch, and Taylor 2004). Hence, it is not surprising that prior research has examined whether developing global or local brands constitutes the optimal strategy when entering foreign markets.

For example, Steenkamp, Batra, and Alden (2003) examine reasons that some consumers prefer global brands to local brands. Their findings suggest that perceived brand globalness might posi-

tively impact purchase likelihood by increasing perceived brand quality and prestige. Similarly, Batra et al. (2000) find that brands perceived as "nonlocal" (from outside the country) in India are preferred to brands perceived as local, for reasons that go beyond brand quality assessment (as has been shown for developed countries). Specifically, their findings suggest that a brand's country of origin stands not only for the summary of a product's quality, but also for the degree of foreignness or nonlocalness, which enhances consumer status and, therefore, brand liking.

International marketing managers face further branding challenges once they have *entered* a foreign market. It is common in both national and foreign markets that companies (e.g., Nike) leverage strong brand equity by extending their brands into other product categories. As a result, one challenge marketing managers face in foreign markets is whether they should extend their brands into related, congruent (e.g., Nike golf shirts), or distant, incongruent (e.g., Nike microwave ovens) product categories.

The domestic literature on brand extensions has developed extensively since the late 1980s and early 1990s. Much of this effort has focused on finding relevant antecedent and moderating variables that affect consumers' brand extension evaluations. Factors analyzed as antecedents to brand extension evaluation include the relationship between the parent brand and its extension (e.g., Aaker and Keller 1990), the parent brand's characteristics (e.g., Dacin and Smith 1994), and the brand extension's product category characteristics (e.g., Smith and Park 1992).

Despite such research, studies involving brand extension strategies in foreign markets are very limited. Indeed, as we will discuss later, there exist only a handful of studies that examine brand extensions in cross-cultural settings. For example, Han and Schmitt (1997) find that perceived fit (company size) is a more important determinant of brand extension evaluation for U.S. (Hong Kong) consumers. In addition, Monga and John (2007) find that the style of thinking (analytic versus holistic) helps explain cultural differences in brand extension evaluations. The fact that so few studies exist limits our understanding of effective brand extension strategy in a cross-cultural context. To facilitate such research, this article proposes a new conceptual framework and several propositions regarding effective brand extension strategy in a cross-cultural context.

Specifically, we propose ways in which culture may affect consumers' brand extension evaluation. In so doing, we develop a research agenda on the cross-cultural brand extension issue by presenting propositions that are in need of empirical validation. As such, the present article does not constitute a full review of factors that affect consumers' brand extension evaluation (for recent brand extension reviews, see Czellar 2003; Hem, de Chernatony, and Iversen 2003; Völckner and Sattler 2006), but rather focuses on selected factors that seem particularly relevant in cross-cultural brand extension theory and research.

In the following, we will first review more commonly examined antecedent variables of (national) brand extension evaluation as discussed in the current marketing literature. We will then propose a definition of culture and subsequently review the existing cross-cultural brand extension research. Next, drawing upon Hofstede's (2001) cultural dimensions and Roth's (1995) "theory of socioeconomics," we will develop propositions that will provide a cross-cultural brand extension conceptual framework for stimulating research about brand extensions across cultures. We will conclude by summarizing our propositions and outlining managerial implications.

Literature Review

Success Factors from Current National Brand Extension Research

Brand extension research has examined many factors that influence national brand extension evaluation. It is not our purpose to review all of these factors in detail, but rather to group them into

broad categories and discuss and present our propositions for selected factors in each group that we believe have been most intensely investigated in prior research. This modus operandi ensures that the propositions we put forward for the moderating influence of cultural dimensions will also hold true for other factors not listed in the respective category because we expect the underlying relationships between the broad category and cultural dimensions to be similar across all specific factors within that category.

Prior brand extension research has repeatedly identified at least three broad determinants of brand extension evaluation (Czellar 2003; Dacin and Smith 1994; Hem, de Chernatony, and Iversen 2003; Keller and Aaker 1992; Park, Jaworski, and MacInnis 1986; Park, Milberg, and Lawson 1991; Smith and Park 1992; Völckner and Sattler 2006). First and foremost, researchers have most frequently investigated the *relationship between the parent brand and the brand extension* as an antecedent, and most important, how *perceived similarity of the extension to the parent brand* affects brand extension evaluation. Perceived similarity, or fit, between the parent brand and its extension has been defined as the degree to which consumers perceive the brand extension to be similar to other products affiliated with the parent brand (Hem, de Chernatony, and Iversen 2003; Smith and Park 1992). As perceived similarity between the parent and the extended brand increases, the consumer's positive evaluation of the extension increases as well (e.g., Aaker and Keller 1990; Boush et al. 1987; Boush and Loken 1991; Dacin and Smith 1994). It has been argued that this finding is due to the positive affect transfer that takes place between the parent brand and its extension when the brand extension is relatively congruent to its parent brand (e.g., Boush and Loken 1991). In contrast, if the brand extension is relatively incongruent to its parent brand, no such affect transfer takes place, leading to less favorable brand extension evaluations.

Within the broader category of the relationship between the parent brand and the brand extension, *brand concept consistency and relevance of the brand's specific association in the extension category* have also been found to affect brand extension evaluation. Regarding the former, Park, Jaworski, and MacInnis (1986) define a brand's concept to refer to the type of need(s) the brand satisfies. Satisfying a performance-related need characterizes a brand with a functional image, whereas helping consumers associate themselves with a desired group, role, or self-image characterizes a brand with a symbolic image. In their normative model, Park, Jaworski, and MacInnis (1986) suggest that in the fortification stage brands be extended into product categories similar to their original brand concept. To illustrate, a brand with a functional concept is suggested to be best extendable to other performance-related product categories. Park, Milberg, and Lawson (1991) find empirical support for this notion of *brand concept consistency.* Specifically, they find that brands with functional brand concepts such as Timex can better extend into functional (e.g., batteries and calculators) than symbolic categories (e.g., bracelets and rings). In contrast, symbolic brands such as Rolex can better extend into symbolic than functional categories.

Similarly, Broniarczyk and Alba (1994) build upon this research stream by demonstrating across several product categories that *brand-specific associations,* which are narrower than Park, Milberg, and Lawson's (1991) functional/symbolic distinction, moderate and even dominate the effect of brand affect and product category similarity on extension evaluation. A brand-specific association is defined "simply as an attribute or benefit that differentiates a brand from competing brands" (Broniarczyk and Alba 1994, p. 215). The authors find that brand-specific associations moderated the role of product category similarity in brand extension judgments such that a brand extension was more preferred in a dissimilar category that valued its association than in a similar category that did not value its association. Overall, therefore, brand concept consistency and relevance of a brand's specific associations in the extension category have been found—together with the

perceived similarity of the extension to the parent brand—to affect consumers' brand extension evaluation within this first broader determinant of brand extension evaluation (the relationship between the parent brand and the brand extension).

Second, prior research has also examined potential effects of several *parent brand characteristics* on brand extension evaluation (Völckner and Sattler 2006). For example, Smith and Park (1992) find that the *perceived quality*, or brand strength, of the parent brand positively affects brand extension evaluation. This notion is based on the reasoning that brands of high quality constitute less risk for consumers and as such are better suited to stimulate trial with less investment than brands of lower quality (Aaker and Keller 1990; Smith and Park 1992). Keller and Aaker (1992) find that high-quality brands stretch further than average-quality brands, supporting the notion that perceived parent brand quality affects brand extension evaluation. Others have further found support for the positive relationship between the level of perceived quality and brand extension evaluation (e.g., Bottomley and Doyle 1996; Dacin and Smith 1994; Sunde and Brodie 1993). Finally, Hem, de Chernatoy, and Iverson (2003) point out that, in the context of brand extension research, brand reputation has been defined in terms of consumers' perception of the parent brand quality.

Another parent brand characteristic that has been shown to affect consumers' brand extension evaluation is *the number of product categories affiliated with a brand.* Specifically, Dacin and Smith (1994) find a positive relationship between the number of product categories affiliated with a brand and consumers' confidence in and favorability of their evaluations of extension quality. In general, it is argued that the more product categories successfully affiliated with a parent brand, the more favorable consumers are toward new brand extensions because previous extensions were successful. As such, consumers use the number of product categories affiliated with a parent brand as a rule of thumb to infer via heuristic decision making their attitude toward a new brand extension (Dacin and Smith 1994). Nisbett, Krantz, and Jepson (1983) find support for this notion by demonstrating that consumers were comparatively more confident about their judgment if it was based upon a comparatively large (versus small) number of instances. Accordingly, it can be argued that "the products affiliated with a brand essentially represent a 'database' from which consumers draw information in forming judgments about an extension. As the number of products increases, the breadth of data as well as the sheer number of data points increases—that is consumers acquire knowledge of the brand in multiple product contexts" (Dacin and Smith 1994, p. 232). This reasoning implies, however, that consumers' confidence in their brand associations as well as their attitude toward a new brand extension should be greater and more favorable, respectively, if they have acquired a great amount of brand knowledge. The opposite holds true if consumers have only obtained little brand knowledge. As a result, the sheer number of products affiliated with a brand constitutes a signal of a given level of brand extension quality.

Along similar lines, the fact that the number of products affiliated with a brand will affect brand extension evaluation can also be derived from taking into consideration the possible negative impact of a poor-quality brand extension introduction. Specifically, it can be argued that a firm with several affiliated products will do everything in its power to avoid the introduction of a poor-quality brand extension because of the negative consequences such a brand extension introduction would have not only on the extension itself but also on all of the other (future revenues expected from) products affiliated with the brand. Again, therefore, consumers view the number of products affiliated with a brand as a signal of the brand extension's probable success.

Third, it has been found that a *brand extension's product category characteristics* affect consumers' brand extension evaluations. Most important, *perceived risk* associated with the brand extension category negatively affects brand extension evaluation (Völckner and Sattler 2006) and positively affects the opportunity to accrue price premiums (DelVecchio and Smith 2005).

This reasoning is based on the view that brand names constitute schemas that hold attributes and beliefs about brand-related experiences (DelVecchio and Smith 2005; Sujan and Bettman 1989). Therefore, consumers use such brand schemas with all of their stored knowledge as heuristics to reduce the perceived risk involved when making purchase decisions (Wernerfelt 1988).

Specifically, DelVecchio and Smith (2005) distinguish between social and financial risk associated with the extension product category. *Social risk* refers to the social consequences when purchasing a product. It "increases to the extent that the product is subject to peer evaluation and is visibly branded" (DelVecchio and Smith 2005, p. 188). Social risk is present if consumers believe that their peers would think less of them for purchasing a given brand extension (Harrell 1986). As such, the aspect of public consumption constitutes a driver of social risk that many companies (e.g., in the apparel and cosmetics industry) are now taking into consideration by positioning their products and brands as "in" to alleviate consumers' perceived social risk (DelVecchio and Smith 2005). Furthermore, social risk exists if the brand extension is readily visible (e.g., athletic shoes) rather than relatively invisible (e.g., a belt). As a result, the extent of public consumption and visibility of a brand determine the level of perceived social risk. Bearden and Etzel's (1982) finding that other people's opinions become more important in a brand selection process if a product is visibly consumed supports this line of reasoning.

In contrast, *financial risk* refers to the economic cost to the consumer if the extension does not perform adequately (DelVecchio and Smith 2005; Grewal, Gottlieb, and Marmorstein 1994). DelVecchio and Smith (2005) find that both social and financial risk affect brand extension price premiums. This finding suggests that the perceived social and financial risk of the brand extension category both also affect consumers' brand extension evaluation. In particular, brand extensions are likely to be evaluated more favorably if the perceived social and financial risk in the extension category are reduced. Völckner and Sattler (2006) find empirical support for this reasoning.

In sum, consumers' brand extension evaluations have been demonstrated to be a function of the relationship between the parent brand and the brand extension, parent brand characteristics, and the brand extension's product category characteristics. As we will argue below, however, cultural dimensions are likely to moderate the effects of these antecedents. Before doing so, however, we propose a definition of culture and review the very limited brand extension research that has been conducted in a cross-cultural context.

The Meaning of Culture

While we have discussed relevant determinants of national brand extension evaluation in the previous section, it remains unclear how these same determinants affect brand extensions in cultures outside of the United States. As a basis for this analysis, however, a common understanding of "culture" is needed. In this section, we therefore develop a definition of culture which serves as the foundation for our further discussion. This is even more important since agreement on a single definition of culture remains elusive.

Definitions of culture have been proposed in anthropology, sociology, cross-cultural psychology, and marketing. Hofstede (2001) has been very influential in advancing cross-cultural research. He defines culture as "the collective programming of the mind which distinguishes the members of one group or category of people from another" (2001, p. 9). Central to this definition is the idea of shared values (in contrast to between-values) held by a society. Furthermore, Hofstede (1991) regards culture as a multilayer system. He uses the metaphor of an onion to suggest that culture is a system with *values* forming the "invisible" core. Subsequent layers are more observable than values and include: *rituals*, *heroes*, and *symbols* and *practices.* Overall, therefore, Hofstede acknowledges that culture has an "invisible" and a "visible" component to it.

Similar to Hofstede's (1991, 2001) model of culture, Hannerz (1992) also distinguishes between an "invisible" and a "visible" dimension of culture with his concept of cultural flow. Specifically, he argues that culture consists of an "invisible" dimension—*ideas and modes of thought*—that encompasses values, beliefs, experiences, meanings, concepts, knowledge, and mental operations that people within some social unit share. In addition, he argues that culture also consists of a "visible" dimension, that is, of *forms of externalization.* This dimension of culture highlights "the different ways in which ideas and modes of thought are made public and accessible to the senses" (Hermans and Kempen 1998, p. 1115; e.g., forms of art and media, interstate highways, food, fashion, music, brands). Moreover, Hannerz (1992) proposes a third dimension of culture, *social distribution*, to take into consideration the distribution of external forms ("visible" component) as well as ideas and modes of thought ("invisible" component) within a given national culture or region. This dimension of culture is crucial because it acknowledges the susceptibility of culture to global dynamics as well as the heterogeneity of subgroups within a population with respect to cultural meanings ("invisible") and practices ("visible"; Hermans and Kempen 1998; Miller 1997). In the words of Hermans and Kempen (1998, p. 1116), "cultural knowledge is to some extent shared, but the degree of sharedness varies. Some subgroups have more access to specific cultural messages than others, partly as a result of differential influence on their formation." A global context with social distribution and, therefore, exposure to different meaning systems and forms of externalization is likely to result in a "dynamic multiplicity" of different positions and voices of one's self and identity (Hermans and Kempen 1998, p. 1118).

In sum, Hofstede (1991, 2001) and Hannerz (1992) both substantiate that culture as a construct encompasses a visible and an invisible component. This line of reasoning is in accordance with Spencer-Oatey (2000, p. 4), who introduces an interpretive element of the cultural concept (Dahl 2004). Accordingly, she defines culture as "a fuzzy set of attitudes, beliefs, behavioural norms, and basic assumptions and values that are shared by a group of people, and that influence each member's behaviour and his/her interpretations of the 'meaning' of other people's behavior." These perspectives all suggest that technologies, symbols, and norms may be mutually understood, but not necessarily shared, by a certain society or culture due to different interpretations of the meaning of other people's behavior.

Taken together, we suggest that the invisible component of culture includes shared values, beliefs, and experiences, which find public expression in the form of symbols and norms. Furthermore, scholars agree that culture can also be visible in many aspects. As such, culture includes a shared set of tools necessary for living, such as food, fashion, and media (collectively referred to as "shared technologies"). However, Hofstede's argumentation appears more restrictive in the context of increasing globalization. Hence, we embrace Hannerz's (1992) argument that the degree of shared norms, symbols, and technologies may vary even within one and the same population (see also Spencer-Oatey 2000, p. 4). In short, norms, symbols, and technologies may be mutually understood, but not necessarily shared, by a certain society or culture due to different interpretations of the meaning of other people's behavior.

Based on this reasoning, we employ the following definition of culture: Culture is a mutually understood set of norms, symbols, and technologies that differentiate one group of people from another and that individuals within a group share to varying degrees (Merz, He, and Alden, forthcoming). Norms and symbols both represent the more "invisible" dimension of culture. Social norms encompass such things as mutually understood behavior rules or social expectations. A certain cultural group uses symbols to express its self and self-identity via participation in specific lifestyle and consumption behaviors, including use of particular brands. Technologies represent the "visible" dimension of culture, under which we subsume all of the tools for living—for example forms of fashion, media, art, food, music, brands, and so forth.

For this definition of culture, it is crucial to understand that all of these components of culture are mutually understood but not necessarily shared to the same degree by a certain cultural group. Therefore, rather than embracing the "metaphor of culture as an internally homogeneous society," we acknowledge the heterogeneity, interconnectedness, complexity, and deterritorialization of culture (Hermans and Kempen 1998). As such, we recognize that "cultural" or "regional" differences could also exist within one nation/country. For our further examination of cross-cultural brand extension evaluation, therefore, we not only examine cultural differences measured by Hofstede's (2001) cultural dimensions (e.g., power distance, uncertainty avoidance, individualism versus collectivism) but also by Roth's (1995) "theory of socioeconomics." Especially the latter consideration takes into account "regional" differences within one nation/country, too, which satisfies the viewpoint of cross-cultural researchers who argue that people within one nation (e.g., U.S. East Coast versus U.S. West Coast) might differ from each other significantly more than people from different nations (e.g., Hong Kong versus New York) in terms of the set of norms, symbols, and technologies used in everyday life.

The Focus of Current Cross-Cultural Brand Extension Research

With our definition of culture as the unit of analysis in mind, we will now review the very limited existing cross-cultural brand extension research. In general, researchers have pointed out that cultural differences may account for different findings in brand extension research that was undertaken outside of the United States. For example, Sunde and Brodie (1993) replicated Aaker and Keller's (1990) seminal brand extension research study in New Zealand and obtained somewhat different results. Aaker and Keller (1993) suggested that cross-cultural factors might have contributed to these differences (e.g., variations in beliefs about the appropriateness of firms introducing multiple products under the same brand name and variations in opinions about the appropriateness of brand extensions for different product categories or parent brands). Such a conclusion points to the importance of considering cultural influences when developing a company's global brand extension strategy. While only limited research has examined consumers' brand extension evaluations in the global marketplace, published studies demonstrate that cross-cultural differences exist.

Han and Schmitt (1997) were the first to examine brand extension success factors across cultures. Specifically, they examined whether consumers' attention would be drawn to characteristics of the company providing the extension (e.g., company size) or to product-category related factors (e.g., the fit between the extension and the core product). Interestingly, they found that perceived fit is much more important for U.S. consumers. In contrast, company size matters more for Hong Kong consumers, but only for low-fit brand extensions. Overall, the authors hypothesize that such differences in brand extension evaluation are due to the collectivistic culture predominant in Hong Kong and the individualistic culture predominant in the United States. They argue that Hong Kong consumers "rely on companies as interdependent, collective societal entities to reduce the risk of a low-fit extension, whereas U.S. consumers—as individualists—place higher importance on their own judgment regarding the product fit rather than cues such as company size" (Han and Schmitt 1997, p. 77).

Bottomley and Holden (2001) undertook a secondary analysis of the original data of eight brand extension studies conducted around the world. They found that the importance of the original brand's quality, the fit between the parent and the extension category, and the interaction of the two on brand extension evaluation varied across cultures. Specifically, they found that cultural differences do not change the fact that the main effects of parent brand quality and the fit between the parent brand's product category and its extension contribute significantly to consumers' brand

extension evaluation. Interestingly, however, they demonstrated that cultural differences do influence the *relative* importance of these factors and their interaction with each other (as measured by standardized regression coefficients). To determine the impact of cross-cultural effects, Bottomley and Holden (2001) performed a series of F tests. Their null hypothesis that the regression coefficients are the same across data sets was rejected in seven of eighteen brand extensions and approached significance in two instances. Overall, their results suggest that the coefficients of the full effects model were not the same across the data sets.

In a different vein, Monga and John (2004, 2007) examined whether culture affects the way consumers evaluate the fit between the brand extension and the parent brand. Specifically, they looked at whether different styles of thinking (U.S. consumers: analytic; Indian consumers: holistic) affect consumer judgment of brand extension fit. Interestingly, these researchers found that consumers from another culture (India) also consider the fit between the parent brand and its extension when evaluating brand extensions. However, cultural differences still existed in terms of brand extension evaluation: Indian consumers judged even poorly fitting extensions (by U.S. norms) as at least moderate fits with the parent brand, which translated to more positive evaluations of such low-fit brand extensions. Moreover, Monga and John (2007) found support for their hypothesis that the style of thinking (analytic versus holistic) is responsible for cultural differences in brand extension evaluations. Specifically, Westerners evaluated brand extensions more favorably under the holistic thinking condition, which resulted in increased perception of the brand extension fit. In contrast, Easterners evaluated brand extensions less favorably under the analytic thinking condition, which resulted in decreased perception of the brand extension fit.

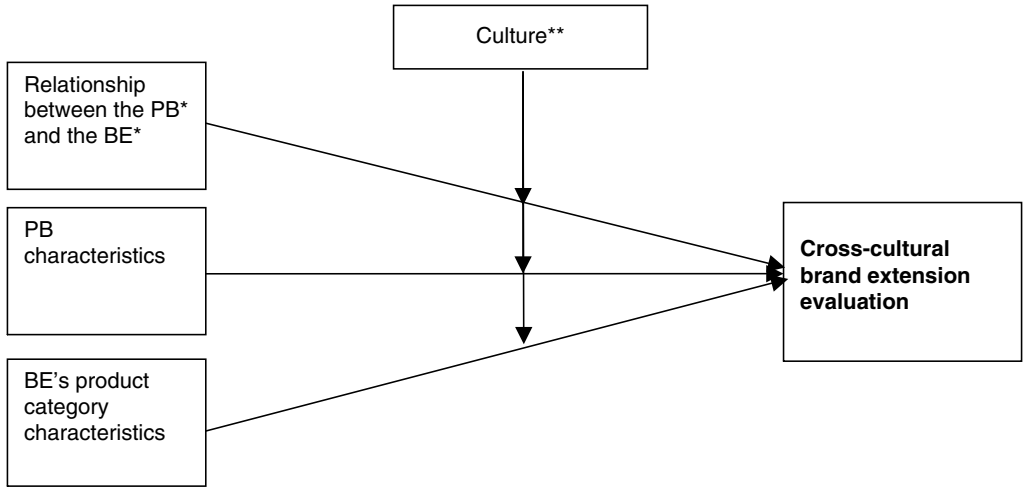
Similarly, Ng and Houston (2004) found that different types of representations are differentially accessible across cultures which affects brand attitude and brand extension evaluations. Specifically, Singaporeans were found to more likely list exemplars (retrieve information about the context, holistic), while Americans were more likely to list beliefs (focus on attributes of the object, analytic) in a free association task. In addition, subjects primed with an independent self-view retrieved more beliefs or attributes about the brand, while subjects primed with an interdependent self-view retrieved more exemplars. Therefore, people with different self-views differ in the information they access in brand extension evaluations.

Finally, Yoon and Gurhan-Canli (2004) report that American and Korean consumers evaluate high-fit brand extensions similarly. However, when the perceived fit was relatively low, Korean (versus American) consumers evaluated the extension more favorably and showed a tendency to incorporate more nondiagnostic information (e.g., negative information about the CEO of the company) into the decision making process.

In sum, very early studies replicated Aaker and Keller's (1993) seminal national brand extension study in other cultures to test whether their results can be generalized across cultures. Only recently have researchers begun to examine *why* cultural differences in brand extension evaluation exist. It has been found that consumers' value systems (e.g., individualism versus collectivism; Han and Schmitt 1997) and their styles of thinking (e.g., Monga and John 2007; Ng and Houston 2004) account for differences in brand extension evaluation.

As a result, while some researchers have begun to conduct brand extension research in a cross-cultural setting, research remains limited. This is surprising, given the risk and financial resources involved in introducing a brand extension (that was successful in the brand's home market) into a foreign marketplace. This risk and financial burden, together with the fact that brand extensions constitute a popular marketing strategy in the global marketplace, increase the importance of understanding cross-cultural differences in brand extension evaluation. In the following section,

Figure 4.1 **A Cross-Cultural Framework for Determining Brand Extension Evaluation**



*PB = Parent Brand; BE = Brand Extension
**Power Distance, Uncertainty Avoidance, Individualism versus Collectivism, Socioeconomics

we will, therefore, discuss how culture may moderate the current findings in brand extension research. Conceptually, we propose a cross-cultural framework for determining brand extension evaluation, as depicted in Figure 4.1. We will develop several propositions that can be viewed as a research agenda for further cross-cultural brand extension research.

Cross-Cultural Brand Extension Evaluation Framework

Prior research has established that cultures differ from each other in terms of various value dimensions (Hofstede 2001; Schwartz 1992, 1994). We will focus on Hofstede's (2001) cultural dimensions to argue that the determinants of brand extension evaluation that we have identified in a previous section are likely to be moderated by such cultural values. Hofstede (2001, p. xix) hypothesizes five main dimensions on which national cultures differ and which reflect basic problems that "any society has to cope with but for which solutions differ." Specifically, he identifies power distance, uncertainty avoidance, individualism versus collectivism, masculinity versus femininity, and long-term versus short-term orientation. Three of these appear particularly relevant to our framework. Prior branding research has consistently found that important cross-cultural differences exist for power distance, uncertainty avoidance, and individualism versus collectivism (e.g., Han and Schmitt 1997; Roth 1995). In addition, Roth's (1995) findings of the effects of socioeconomics on the performance of global brand image strategy imply that socioeconomics will also affect brand extension evaluation. This moderating variable is important because it implies that regions within countries may differ from each other as well. As such, it is in line with our acknowledgment of the heterogeneity, complexity, interconnectedness, and deterritorialization of culture (see above; also Hermans and Kempen 1998).

In the following sections, we will briefly describe the three cultural dimensions that are of particular relevance to our analysis (Hofstede 2001), as well as the "dimension" of socioeconomics (Roth 1995), and discuss ways in which they moderate the domestically identified antecedents to

brand extension evaluation previously described. As a result, we present several propositions that we hope will provide a research agenda for future theory-based research.

Power Distance

Power distance describes the extent of cultural inequality that underlies the functioning of each particular society. Cultures high in power distance tend to focus on social status, class affiliation (rich versus poor), and wealth. People of such cultures are motivated to maintain, increase, and show their "power" or social status as a source of satisfaction (Roth 1995). In high power distance cultures, people are particularly motivated to conform to the behavior of other people in their class or people of other classes to which they aspire to belong. Social consciousness is high. In contrast, cultures low in power distance tend to emphasize much less social and economic class affiliation. Such cultures de-emphasize the differences between their citizens' wealth and power (Hofstede 2001).

If people in high (versus low) power distance cultures are relatively more motivated to emphasize the existent inequality in society and belongingness to a particular social group, then it may be that such consumers also more strongly adhere to usage of certain brands. It has been shown that brands constitute important signs that people buy in order to demonstrate social status and class affiliation (Belk, Bahn, and Mayer 1982; Escalas and Bettman 2003). Thus, in high (versus low) power distance cultures, where the consumption of and belongingness to a certain brand is important for social status, it can be hypothesized that consumers will evaluate unrelated (incongruent) brand extensions less favorably.

This reasoning is based on the assumption that affect transfer from the parent brand to the brand extension is more probable when the brand extension is relatively congruent with the parent (Boush and Loken 1991). High power distance consumers appear more likely to be critical of incongruent brand extensions, as they feel anxious about a "breakdown" of categorical boundaries and possible loss of social prestige. In contrast, if the brand extension is very similar to the parent brand, the categorical order should be seen as "preserved" and, therefore, positive affect should transfer more easily from parent to extension.

The literature on feedback effects associated with brand extensions provides a further basis for our reasoning. Prior research has demonstrated that brand extensions can dilute beliefs about a parent brand, particularly when the extension is perceived as moderately incongruent with consumers' expectations for the brand (John, Loken, and Joiner 1998; Loken and John 1993; Milberg, Park, and McCarthy 1997). For example, Milberg, Park, and McCarthy (1997) find that negative feedback effects from the extension to the parent brand occur when the brand extension's product category is perceived as too incongruent and unrelated to the parent brand's product category and when the extension attribute information is too incongruent with image beliefs associated with the parent brand. Relating these findings to power distance, the desire of high power distance consumers to maintain their power distance seems likely to make them more critical of incongruent versus congruent brand extensions (due to a possible loss of social acceptance). Extending this thinking, it can be argued that the desire of high power distance consumers for belongingness and social status makes perceived similarity of the parent brand and its extension a stronger determinant of brand extension evaluation than the desire to acquire a particular brand for the sake of social status and group affiliation. Hence, the danger exists that the extension will dilute the parent brand and vice versa. In cultures low in power distance, these considerations appear to play a less significant role because the maintenance of hierarchic and other categorical boundaries tend not to be as primary a goal to members. In sum, *perceived similarity* is likely to be a more important determinant of brand extension evaluation in cultures high on power distance.

P1 (Perceived similarity): Perceived similarity between the parent brand and its extension constitutes a more important determinant of brand extension evaluation in cultures high in power distance than in cultures low in power distance.

The above reasoning also suggests similar effects for *brand concept consistency* (and brand specific associations).

P2a (Brand concept consistency): Perceived brand concept consistency between the parent brand and its extension constitutes a more important determinant of brand extension evaluation in cultures high in power distance than in cultures low in power distance.

We next consider how cultures varying on power distance respond to congruent (versus incongruent) extensions of brands with functional versus symbolic brand concepts. Prior research has shown that symbolic (or prestige) brand image strategies enhance product performance in high power distance cultures but hurt market share in low power distance cultures (Roth 1995). Again, brands constitute important signs of social status and class affiliation (Belk, Bahn, and Mayer 1982; Escalas and Bettman 2003). Therefore, consumers in high (versus low) power distance cultures will favor brand extensions that satisfy symbolic consumer needs over ones that satisfy functional consumer needs. However, this relationship is likely to be moderated by the level of congruency between the parent brand and its extension such that congruent brand extensions of brands with a symbolic brand concept will be more favorably evaluated than congruent brand extensions of brands with a functional brand concept.

Since extensions vary in their level of congruity with the parent brand concept, how would cultures differing on power distance respond to incongruent extensions? We posit that consumers in high (versus low) power distance cultures will favor incongruent brand extensions of parent brands with a symbolic brand concept less than if such incongruity was demonstrated by brand extensions of parent brands with a functional brand concept. This is because the associations from the parent brand are less likely to be transferred to the incongruent extension which would prevent the extension from delivering its social status and class affiliation associations to consumers. In contrast, consumers in low power distance cultures, where social and economic affiliation is less emphasized, will pay less attention to the respective brand concept when evaluating a brand extension. Overall, therefore, it is likely that brand concept consistency (and a brand's specific associations) constitutes a more important driver of brand extension evaluation in cultures high (versus low) in power distance.

P2b (Brand concept consistency—congruent extensions): In high (versus low) power distance cultures, *congruent* extensions of brands with a symbolic brand concept will be *more* favorably evaluated compared to congruent extensions of brands with functional brand concepts.

P2c (Brand concept consistency—incongruent extensions): In high (versus low) power distance cultures, *incongruent* extensions of brands with a symbolic brand concept will be *less* favorably evaluated compared to incongruent extensions of brands with functional concepts.

Next, we discuss propositions pertaining to two factors (perceived quality and number of affiliated products) of our second identified broader category of brand extension evaluation (parent brand characteristics). The previous discussion suggests further the importance of the *parent brand's*

perceived quality for consumers in high power distance cultures. Again, people in such cultures focus on social status and class affiliation. Therefore, congruent brand extensions from parent brands known for quality constitute a less risky choice for all consumers. In contrast, potential ownership of an incongruent brand extension even from a high-quality parent brand could cause one to lose status in the social hierarchy in the event that affect and knowledge transfer from the parent brand to its extension is hampered. Furthermore, as elaborated, incongruent brand extensions dilute a parent brand's value, thereby endangering high power distance culture consumers' objective to maintain their power distance for social status, prestige, and group affiliation reasons. Overall, therefore, high power distance culture consumers are likely to evaluate congruent (versus incongruent) brand extensions of high-quality parent brands more favorably. On the other hand, consumers from low power distance cultures are likely to assign less weight to the level of a brand extension's incongruency when the parent brand is of high quality, due to less emphasis on affiliation and hierarchic belongingness. As a result, in high (versus low) power distance cultures, the extensions of parent brands with higher perceived quality should be favored relatively more than the extensions of lower perceived quality parent brands.

P3 (Perceived quality): Perceived quality of the parent brand constitutes a more important determinant of brand extension evaluation in cultures high in power distance than in cultures low in power distance.

Moreover, a possible three-way interaction between a parent brand's perceived quality x perceived similarity between the parent brand and its extension x power distance can be proposed. The idea behind this interaction is that many times managers encounter situations in which either the parent brand is not of top quality or the brand is not being extended into a highly similar product category. Under these circumstances, we expect that both high and low power distance consumers will evaluate the following extensions similarly: First, brand extensions whose parent brands are perceived as low-quality and which have only very few product attributes in common with their parent brands (physical similarity) are likely to be evaluated poorly. Second, brand extensions with perceived high-quality parent brands and high physical similarity will be evaluated favorably.

However, we expect that high power distance consumers will evaluate less similar brand extensions from perceived high-quality parent brands more favorably than more similar extensions from poor perceived quality parent brands. This reasoning is based on the assumption that top quality is usually positively correlated with price. Thus, consuming high- versus low-quality products is likely to help high power distance consumers differentiate themselves from others. No such relationship is expected for extensions varying in similarity. In contrast, low power distance consumers will evaluate both extensions similarly.

P4 (Perceived quality x perceived similarity): High power distance consumers will evaluate low-fit brand extensions of high-quality parent brands more favorably than high-fit brand extensions of low-quality parent brands. In contrast, low power distance consumers will evaluate such extensions similarly in terms of their favorableness.

In terms of *number of affiliated products*—our second factor of brand extension evaluation within the broader category of parent brand characteristics—a two-way interaction with power distance is not unique, that is, it could go in two different ways (e.g., number of products affiliated with a brand constitutes a more important determinant of brand extension evaluation in cultures high/low in power distance than in cultures low/high in power distance). Instead, we therefore,

propose a three-way interaction between number of affiliated products x brand concept consistency x power distance.

Specifically, the number of products affiliated with a brand may be viewed as a sign of good quality, or reputation, of the parent brand, as long as those distinct (extension) products are consistent in their concepts with respect to their symbolic parent brand. If this were the case, it can be argued—just as we did above—that brand extensions of such brands constitute a less risky choice for people in high power distance cultures. The likelihood to engage in brand consumption behavior that might negatively affect consumers' position in society will be reduced to a minimum.

In contrast, however, it is also possible that the number of extension products affiliated with a brand may not stand for a particular social class or allow consumers to differentiate themselves from others. In this case, it is likely that a range of people from different social classes consume this brand's products. Hence, the extension products in this case suffer from poor concept consistency. If this were the case, consumers in high power distance cultures are likely to be less affected by the number of products affiliated with a brand in their brand extension evaluations than consumers in low power distance cultures. Thus, we propose the following three-way interaction:

P5 (Number of products affiliated with a brand x brand concept consistency): The number of products affiliated with a symbolic brand constitutes a more important determinant of brand extension evaluation in cultures high in power distance than in cultures low in power distance as long as the concept consistency of these (past) extension products of the symbolic parent brand is high.

Finally, in contrast to people in low power distance cultures, people in high power distance cultures are not only concerned with the parent brand reputation (or with parent brand characteristics in general) as a possible predictor for a brand extension's success but also, and more specifically, with the *social and financial risk associated with the new brand extension's product category* (and hence with the brand extension's product category characteristics in more general). High power distance consumers are likely to evaluate brand extensions more favorably when the extension will not cause danger to their social position in society due to class members' negative evaluation (social risk) or due to the brand extension's economic cost if it does not perform adequately (financial risk).

While social risk is the more important issue here, financial risk can also threaten the social position of high power distance consumers. A very expensive but poorly performing brand extension that was bought for prestige purposes might cause members of high power distance cultures to lose status. Furthermore, the money invested in such malfunctioning prestige brand extensions constitutes opportunity costs to such consumers, as other investments in one's social position could have been made instead. In contrast, consumers in cultures low in power distance are likely to be less concerned with the social and financial risk involved in using a certain brand extension due to the deemphasis of inequality in such cultures.

P6 (Social risk): Perceived social risk of the brand extension's product category constitutes a more important determinant of brand extension evaluation in cultures high in power distance than in cultures low in power distance.

P7 (Financial risk): Perceived financial risk of the brand extension's product category constitutes a more important determinant of brand extension evaluation in cultures high in power distance than in cultures low in power distance.

Uncertainty Avoidance

Uncertainty avoidance describes "the extent to which a culture programs its members to feel either uncomfortable or comfortable in unstructured situations" (Hofstede 2001, p. xix). It focuses on the level of tolerance for ambiguity and uncertainty within the society. Accordingly, it describes the extent to which a culture values predictability, stability, and low stress rather than change, new experiences, surprise, and novelty. Cultures high in uncertainty avoidance tend to be less tolerant of ambiguity, more risk averse, and more resistant to change than low uncertainty avoidance cultures (Roth 1995). Members of such cultures tend to avoid uncertainty, while seeking certainty. This creates a rule-oriented society that institutes laws, regulations, rules, and controls with the purpose of reducing perceived uncertainty. The opposite is true for cultures low in uncertainty avoidance. Such cultures are more tolerant of a variety of opinions and less concerned about ambiguity and uncertainty. Moreover, they take more and greater risks and more readily accept change.

Brand extensions that are relatively unrelated to their parent brands (distant or incongruent brand extensions) contain more uncertainty and ambiguity than brand extensions that are closely related to their parent brands (close or congruent brand extensions). This is due to the likelihood that consumers will have greater difficulty relating positive product and brand attributes associated with the parent brand to more distant brand extensions. Because consumers in high (versus low) uncertainty avoidance cultures are less tolerant of ambiguity and uncertainty, it is likely that they will evaluate more distant brand extensions less favorably. In contrast, consumers in low uncertainty avoidance cultures are less risk averse and more open to ambiguous situations that contain uncertainty. Overall, therefore, *perceived similarity* seems to play a more important role in cultures with high uncertainty avoidance than in cultures with low uncertainty avoidance.

P8a (Perceived similarity): Perceived similarity between the parent brand and its extension constitutes a more important determinant of brand extension evaluation in cultures high in uncertainty avoidance than in cultures low in uncertainty avoidance.

Moreover, factors that are likely to reduce consumers' perceived uncertainty or ambiguity in respect to a particular brand extension (e.g., familiarity with the parent brand or targeted advertising efforts) will have a more significant effect on consumers from high uncertainty avoidance cultures due to their higher sensitivity level. As such, it can be hypothesized that consumers from high uncertainty avoidance cultures prefer brand extensions of familiar parent brands over ones of less familiar or less established parent brands. Similarly, targeted advertising has been demonstrated to positively affect consumers' perceived fit between the parent brand and its extension (Lane 2000). Thus, it can be expected that targeted advertising will increase the perceived similarity between the parent brand and its extension (and hence its evaluation), but more so in the case of high (versus low) uncertainty avoidance cultures.

P8b (Perceived similarity): In high (versus low) uncertainty avoidance cultures, brand extensions of more familiar and established parent brands will be evaluated more favorably than of less familiar and established parent brands.

P8c (Perceived similarity): Targeted advertising will have a greater positive effect on brand extension evaluation in high uncertainty than low uncertainty avoidance cultures.

In addition to our argument that distant or incongruent brand extensions contain more uncertainty and ambiguity than close or congruent brand extensions, an even more specific stand can be taken

as a basis for the examination of the relationship between brand extension evaluation and level of incongruity. Research on categorization and schema incongruity has found a nonmonotonic relationship between the level of (in)congruity and elaboration/evaluation (e.g., Mandler 1982; Meyers-Levy and Tybout 1989). Specifically, it has been found that moderately incongruent situations trigger more elaboration and favorable evaluations than both congruent and extremely incongruent situations (for a discussion, see Stayman, Alden, and Smith 1992). This inverted-U relationship between level of congruity and elaboration/evaluation has also been demonstrated in the national brand extension literature (Barone, Miniard, and Romeo 2000; Merz and Alden 2007). Moreover, further support for this nonmonotonic relationship between (in)congruity and elaboration effort can also be derived from the categorization-process and conflict-theory literatures (see, e.g., Ozanne, Brucks, and Grewal 1992).

Overall, therefore, it can be derived from these streams of research that moderately incongruent brand extensions contain the highest level of ambiguity and uncertainty in comparison to both congruent and extremely incongruent brand extensions. Greater elaboration in moderate incongruity conditions is to resolve the ambiguity and its successful resolution results in favorable evaluations. An extremely incongruent brand extension is viewed as too unrelated to the parent brand to attempt solving the resulting ambiguity and uncertainty. As a result, consumers in high uncertainty avoidance cultures are likely to evaluate moderately incongruent brand extensions less favorably than congruent and extremely incongruent brand extensions and less favorably than consumers in low uncertainty avoidance cultures. Furthermore, because of their resolving a challenging ambiguity generated by moderately incongruent extensions, low (versus high) uncertainty avoidance consumers will evaluate moderately incongruent brand extensions more favorably than either congruent or extremely incongruent brand extensions. No difference in brand extension evaluation between high versus low uncertainty avoidance consumers is expected for congruent and extremely incongruent extension.

P8d (Perceived similarity): Moderately incongruent brand extensions will be evaluated more favorably in low uncertainty than high uncertainty avoidance cultures. Congruent and extremely incongruent extensions will be similarly evaluated in low and high uncertainty avoidance cultures.

P8e (Perceived similarity): In low (versus high) uncertainty avoidance cultures, moderately incongruent brand extensions will be evaluated more favorably than both congruent and extremely incongruent brand extensions.

Our second determinant of brand extension evaluation within the broader category of the relationship between the parent brand and the brand extension is the *brand concept consistency* or brand specific associations. Going back to our more general analysis of congruent versus incongruent brand extensions, we have previously argued that incongruent brand extensions contain more uncertainty and ambiguity than congruent brand extensions. Therefore, incongruent brand extensions in general will be evaluated more negatively in high uncertainty avoidance cultures. Furthermore, consumers in high uncertainty avoidance cultures are more focused on using brands for solving and preventing problems (Roth 1995). In contrast, risk aversion is diminished in low uncertainty avoidance cultures, leading to more openness for symbolic and experiential consumption. Therefore, we propose that high uncertainty avoidance consumers will prefer brand extensions (both congruent and incongruent) that are being associated with satisfying functional needs over brand extensions that are being associated with satisfying symbolic needs. More flexibility in terms of brand extension preference, however, can be expected from low uncertainty avoidance consumers.

Overall, therefore, it can be argued that brand concept consistency constitutes a more important determinant of brand extension evaluation in cultures high in uncertainty avoidance.

P9a (Brand concept consistency): In high (versus low) uncertainty avoidance cultures, *congruent* and *incongruent* extensions of brands with a symbolic brand concept will be evaluated *less* favorably compared to similar extensions of brands with functional brand concepts.

P9b (Brand concept consistency): Perceived brand concept consistency between the parent brand and its extension constitutes a more important determinant of brand extension evaluation in cultures high in uncertainty avoidance than in cultures low in uncertainty avoidance.

In terms of parent brand characteristics, the *perceived quality of the parent brand* and *the number of products affiliated with the parent brand* are likely to reduce consumers' perceived risk and increase confidence in brand associations. This, however, seems to be more relevant in cultures high in uncertainty avoidance, where people actively try to avoid uncertainty, than in cultures that are low on uncertainty avoidance, where people are clearly less risk averse and more tolerant of uncertainty. As a result, both the quality of the parent brand and the number of products affiliated with the parent brand are likely to be more important determinants of brand extension evaluation in high (versus low) uncertainty avoidance cultures. In terms of parent brand quality, Erdem, Swait, and Valenzuela's (2006) finding that uncertainty avoidance amplifies the net effect of brand credibility on consumers' choice substantiates our reasoning.

P10 (Perceived quality): Perceived quality of the parent brand constitutes a more important determinant of brand extension evaluation in cultures high in uncertainty avoidance than in cultures low in uncertainty avoidance.

P11 (Number of products affiliated with a brand): The number of products affiliated with a brand constitutes a more important determinant of brand extension evaluation in cultures high in uncertainty avoidance than in cultures low in uncertainty avoidance.

In addition to these two-way interactions, we can further propose three-way interactions for both parent brand characteristics, similar to the ones proposed earlier for power distance. First, a three-way interaction between perceived quality x perceived similarity x uncertainty avoidance can be proposed. We expect no cultural differences between high versus low uncertainty avoidance customers in terms of brand extension evaluation for close (distant) brand extensions of high (low) quality parent brands. Such extensions are likely to be evaluated favorably (unfavorably) by both high and low uncertainty avoidance consumers because close (distant) extensions and high (low) parent brand quality aggregately reduce (increase) consumers' perceived ambiguity, risk, and uncertainty independent of the respective level of uncertainty avoidance.

However, we expect cultural differences for close (distant) brand extensions of low (high) quality parent brands. Prior research has found that perceived fit constitutes a relatively more important factor of brand extension evaluation than parent brand quality (Völckner and Sattler 2006). Accordingly, it can be argued that a poor fit between the parent brand and its extension will result in relatively more uncertainty and ambiguity than a poor perception of a parent brand's quality. This should be particularly the case for high uncertainty avoidance consumers who tend to avoid uncertainty and be less tolerant of ambiguity and novelty. As a result, we argue that high uncertainty avoidance consumers will evaluate high-fit brand extensions of low-quality parent brands more favorably than low-fit

brand extensions of high-quality parent brands. No such difference is expected for low uncertainty avoidance cultures, which are more tolerant and less concerned about ambiguity and uncertainty.

P12 (Perceived quality x perceived similarity): High uncertainty avoidance consumers will evaluate low-fit brand extensions of high-quality parent brands more favorably than high-fit brand extensions of low-quality parent brands. In contrast, low uncertainty avoidance consumers will evaluate such extensions similarly in terms of their favorableness.

Second, a three-way interaction between number of products affiliated with a brand x brand concept consistency x uncertainty avoidance can be proposed. The number of affiliated products is likely to reduce consumers' perceived risk and uncertainty, as long as a brand's concept is kept consistent. As such, the number of products affiliated with a brand constitutes a signal for especially high uncertainty avoidance consumers for the potential success of a new brand extension. However, if the affiliated products do not show a consistent brand concept, then high (versus low) uncertainty avoidance consumers are less affected by the number of affiliated products as a result of the existing ambiguity and uncertainty. In contrast, low uncertainty avoidance consumers are expected to be less affected by the number of affiliated products and the brand concept consistency, due to their lesser sensitivity level with respect to ambiguity and uncertainty.

P13 (Number of products affiliated with a brand x brand concept consistency): The number of products affiliated with a brand constitutes a more important determinant of brand extension evaluation in high uncertainty avoidance cultures than in low uncertainty avoidance cultures as long as the concept consistency of these (past) extension products of the parent brand is high.

Finally, in terms of the brand extension's product category characteristics, the *category's perceived social and financial risk* are likely to constitute a more important determinant in cultures high (versus low) in uncertainty avoidance. This postulation is based on the reasoning that consumers in high uncertainty avoidance cultures aim at avoiding any kind of risk, whether it is the risk of being associated with the "wrong" consumption group (social risk) or the risk of experiencing an economic loss (financial risk). Specifically, high uncertainty avoidance consumers are more likely to avoid the trial of brand extensions that are introduced in product categories that are perceived to have high social risk, due to their avoidance of public embarrassment, and high financial risk, due to the possible economic loss if the extension does not perform adequately. Consumers in low uncertainty avoidance cultures are more risk taking and hence more likely to try a newly introduced brand extension, whether or not the extension's product category is perceived as socially or financially risky.

P14 (Social risk): Perceived social risk of the brand extension's product category constitutes a more important determinant of brand extension evaluation in cultures high in uncertainty avoidance than in cultures low in uncertainty avoidance.

P15 (Financial risk): Perceived financial risk of the brand extension's product category constitutes a more important determinant of brand extension evaluation in cultures high in uncertainty avoidance than in cultures low in uncertainty avoidance.

Individualism versus Collectivism

Hofstede's (2001, p. xx) individualism/collectivism dimension describes the extent to which "individuals are supposed to look after themselves or remain integrated into groups, usually around the

family." As such, individualism versus collectivism describes the degree to which a culture values independence, freedom, and personal and individual time. In other words, this cultural dimension focuses on the extent to which culture reinforces individual or collective achievement and interpersonal relationships. Consumers in individualistic cultures tend to be self-centered, independent from others, and value personal goals. People in such cultures tend to form a larger number of looser relationships. In contrast, consumers in collectivist cultures tend to be more group focused and interdependent. Conformity and group behavior are important. Also important are extended families and collectives in which members take responsibility for other group members.

In terms of brand extensions, collectivistic cultures tend to think more holistically than individualistic cultures (Chiu 1972; Ji, Peng, and Nisbett 2000; Nisbett et al. 2001) and, therefore, have been argued to be more tolerant toward incongruent brand extensions. For this reason, consumers in collectivistic cultures tend to take related information into consideration and evaluate brand extensions within their respective context rather than in isolation. As a result, the perceived similarity between the parent brand and its extension should be less important to consumers' brand extension evaluation in collectivistic versus individualistic cultures. Indeed, as noted earlier, prior cross-cultural brand extension research finds support for this reasoning (Han and Schmitt 1997; Monga and John 2004, 2007; Ng and Houston 2004; Yoon and Gurhan-Canli 2004). In contrast to these findings, however, one study exists that demonstrates that collectivistic consumers are "tougher" in categorizing and have more distinct subcategories compared to individualistic consumers (Jain, Desai, and Mao 2007). According to the implications of this study's findings, therefore, collectivistic consumers are likely to tolerate incongruent brand extensions less than individualistic consumers. Until future research empirically addresses this discrepancy, based on the greater evidence for the former relationship, we propose the following:

P16 (Perceived similarity): Perceived similarity between the parent brand and its extension constitutes a more important determinant of brand extension evaluation in individualistic cultures than in collectivistic cultures.

With respect to the second determinant of brand extension evaluation within the broader category of the relationship between the parent brand and its extension—*brand concept consistency*—we introduce another brand concept that has been found to play an important role in cultures high in individualism: the experiential or sensory brand concept (Roth 1995). An experiential brand concept highlights novelty, sensory gratification, and variety seeking. As previously argued, people in high individualism cultures make decisions, form relationships, and initiate behavior independent of others. Furthermore, they tend to seek hedonistic experiences and variety. Such consumers are likely to favor experiential brand images that emphasize variety, novelty, and individual gratification (Roth 1995). Prior research has found support for this notion. Specifically, Roth (1995) demonstrated that experiential brand images have a positive impact on market share when cultural individualism is high. Interestingly, it was also found that functional brand images have a positive impact on market share in high individualism cultures (Roth 1995).

In contrast, for collectivistic consumers, group behavior, conformity, and collective thinking and acting are important. Symbolic brand concepts tend to reinforce group membership and affiliation. Thus, low individualism consumers are likely to value symbolic brand concepts. Indeed, Roth (1995) found support for this notion by discovering that symbolic brand concepts have the greatest impact on market share when cultural individualism is low. Overall, therefore, we expect that collectivistic consumers will evaluate congruent brand extensions of brands with a symbolic brand concept more favorably compared to congruent brand extensions with a functional or experiential

brand concept. In contrast, we expect that individualistic consumers will evaluate congruent brand extensions of brands with an experiential or functional brand concept more favorably compared to congruent brand extensions with a symbolic brand concept.

P17a (Brand concept consistency): In collectivistic (individualistic) cultures, *congruent* extensions of brands with a symbolic (experiential or functional) brand concept will be evaluated more favorably compared to congruent extensions of brands with functional or experiential (symbolic) brand concepts.

Taking into consideration the more holistic approach of collectivistic consumers and the more analytical approach of individualistic consumers in brand extension evaluations, it can further be argued that collectivistic consumers will be more flexible and tolerant than individualistic consumers when it comes to the evaluation of incongruent brand extensions that have the same brand concept as the parent brand. Therefore, it can be argued that collectivistic consumers will evaluate distant, or incongruent, brand extensions of symbolic parent brands more favorably than individualistic consumers as long as brand concept consistency exists. Because individualistic consumers are less flexible and tolerant with respect to the evaluation of incongruent brand extensions, it is unclear how they would evaluate incongruent brand extensions of experiential or functional parent brands when the brand concept is kept consistent. Therefore, we only propose the following:

P17b (Brand concept consistency): In collectivistic (versus individualistic) cultures, *incongruent* extensions of brands with a symbolic brand concept will be evaluated more favorably than incongruent extensions of brands with functional or experiential brand concepts.

Overall, because of the collectivistic (versus individualistic) consumers' more holistic thinking, and hence increased flexibility and tolerance with respect to incongruent brand extensions and brand concept inconsistencies, we expect that brand concept consistency will be a more important determinant of brand extension evaluation in individualistic than in collectivistic cultures.

P17c (Brand concept consistency): Perceived brand concept consistency between the parent brand and its extension constitutes a more important determinant of brand extension evaluation in individualistic cultures than in collectivistic cultures.

Prior cross-cultural brand extension research has found that collectivistic consumers take a more holistic approach to brand extension evaluation than individualistic consumers (Monga and John 2004, 2007; Ng and Houston 2004). Similarly, collectivistic consumers have been found to incorporate nondiagnostic cues (e.g., information about the CEO of a company—Yoon and Gurhan-Canli 2004, or the company size—Han and Schmitt 1997) into their evaluations of incongruent brand extensions. Individualistic consumers do not, focusing instead on perceived similarities between the parent brand and its extension. However, no cultural differences were found for the evaluation of congruent brand extensions (Yoon and Gurhan-Canli 2004). In addition, Han and Schmitt (1997) found that collectivist consumers took company size into consideration when evaluating brand extensions, but only for incongruent brand extensions.

Overall, these findings suggest that individualistic consumers predominantly rely on perceived similarity between the parent brand and its extension when evaluating both congruent and incongruent brand extensions. In contrast, collectivistic consumers are likely to rely primarily on the perceived fit between the parent brand and its extension when evaluating congruent brand exten-

sions. Affect and knowledge transfer from the parent brand to the extension in this case can easily take place, thus minimizing the possibility of losing face and/or deviating from in-group norms.

However, incongruent brand extensions may not only hamper affect and knowledge transfer but also potentially dilute the parent brand. In this case, therefore, collectivistic consumers—who tend to subordinate their personal interests and choices to an existing in-group norm—are more likely to look for, and incorporate into their decision, a parent brand's perceived quality, or reputation, as a more nondiagnostic cue for the appropriateness of such incongruent extensions. This reasoning is in line with the argument that brands in collectivistic cultures are often used to emphasize and reinforce group membership and affiliation. Therefore, we expect that perceived quality, or reputation, of the parent brand constitutes a more important predictor of brand extension evaluation in collectivistic than in individualistic cultures.

P18 (Perceived quality): Perceived quality of the parent brand constitutes a more important determinant of brand extension evaluation in collectivistic cultures than in individualistic cultures.

Again, we can propose a three-way interaction between perceived quality x perceived similarity x individualism/collectivism. Specifically, we expect that both individualistic and collectivistic consumers will evaluate brand extensions that are similar to their high- quality parent brand favorably. In contrast, we expect that both individualistic and collectivistic consumers will evaluate brand extensions that are dissimilar to their low-quality parent brand unfavorably. However, we expect that cultural differences exist for brand extensions that are high on one determinant of brand extension evaluation (e.g., perceived quality) but low on the other (e.g., perceived similarity). Specifically, it is very important for collectivistic consumers to show their belongingness and affiliation to an in-group and to distinguish themselves from other groups (out-group). High-quality brands are often used to reinforce consumers' belongingness to their in-group. In contrast, low-quality brands are seen as carrying the potential danger of losing face. Therefore, we can expect that perceived quality of the parent brand constitutes a more important determinant of brand extension evaluation for collectivistic consumers than the perceived similarity between the parent brand and its extension. Our elaboration above about the more holistic (versus analytic) thinking of collectivistic consumers strengthens this argument. Accordingly, we expect that low-fit brand extensions of high-quality parent brands will be evaluated more favorably by collectivist consumers than high-fit brand extensions of low-quality parent brands. Because individualistic consumers are more independent and value personal freedom, no such difference in brand extension evaluation is expected.

P19 (Perceived quality x perceived similarity): Collectivistic consumers will evaluate low-fit brand extensions of high-quality parent brands more favorably than high-fit brand extensions of low-quality parent brands. In contrast, individualistic consumers will evaluate such extensions similarly in terms of their favorableness.

Next, we propose an interaction effect between the number of products affiliated with the parent brand and individualism/collectivism interaction effect. Consumers in collectivistic cultures are likely to also take nondiagnostic information into consideration in order to ensure that the brand extension is consistent with social and group norms (holistic thinking). Indeed, as noted previously, the findings of prior cross-cultural brand extension research provide support for this hypothesis. Not only has prior research found that consumers in East Asian cultures (predominantly high in

collectivism) incorporate nondiagnostic information, such as information about the CEO of a company, into their brand extension evaluation (Yoon and Gurhan-Canli 2004), but also that company size matters more for consumers in collectivistic cultures than for consumers in individualistic cultures (Han and Schmitt 1997). Assuming that company size is partly a function of the number of products affiliated with a brand, it can be hypothesized that the number of products affiliated with the parent brand will affect brand extension evaluations of consumers in high collectivistic cultures, and significantly more so than it will consumers in high individualistic cultures.

P20 (Number of products affiliated with a brand): The number of products affiliated with a brand constitutes a more important determinant of brand extension evaluation in collectivistic cultures than in individualistic cultures.

Again, we can propose a three-way interaction between the number of products affiliated with a brand x brand concept consistency x individualism/collectivism. As noted, collectivistic consumers care about what others say and think. Moreover, they use brands to emphasize group membership and affiliation. As such, brands—in particular symbolic brands—constitute a means to gaining and saving face.

Furthermore, collectivistic consumers use the number of affiliated products as nondiagnostic information to ensure that the brand extension is consistent with the expected social and group norms. For this purpose, however, it is important that the (symbolic) brand concepts of the products affiliated with the parent brand are all consistent with the parent brand's (symbolic) concept, to avoid confusion and potential loss of face. In this case, the potential danger of brand consumption with negative consequences for one's in- and out-group will be reduced to a minimum. In contrast, however, if the affiliated products have different brand concepts, they are likely to play a less significant role—as nondiagnostic information—in collectivistic than individualistic consumers' brand extension evaluation.

P21 (Number of products affiliated with a brand x brand concept consistency): The number of products affiliated with a brand constitutes a more important determinant of brand extension evaluation in collectivistic cultures than in individualistic cultures as long as the concept consistency of these (past) extension products of the parent brand is high.

Last but not least, the cultural dimension of individualism versus collectivism is also likely to moderate the relationship between a brand extension's product category characteristic and consumers' brand extension evaluation. On the one hand, perceived social risk of the extension category is likely to matter more to consumers in high collectivistic than individualistic cultures. Again, collectivistic (versus individualistic) consumers are more concerned with saving and gaining face and will publicly emphasize group membership and affiliation. As a result, social risk associated with a brand extension's product category will affect brand extension evaluation more in collectivistic than in individualistic cultures, due to a possible loss of face.

On the other hand, the fact that people in high individualistic cultures value individual achievement and goals and that people in high collectivistic cultures value group achievement and goals suggests that financial risk associated with the brand extension's product category constitutes a more important determinant of brand extension evaluation in individualistic than collectivistic cultures. Prior research about the self and risk-taking provides support for this reasoning. For example, Mandel (2003) finds that consumers primed on interdependence (predominantly the case in collectivistic cultures) were less likely to take social risk but more likely to take financial risk than were those primed on independence (predominantly the case in individualistic cultures).

P22 (Social risk): Perceived social risk of the brand extension's product category constitutes a more important determinant of brand extension evaluation in collectivistic cultures than in individualistic cultures.

P23 (Financial risk): Perceived financial risk of the brand extension's product category constitutes a more important determinant of brand extension evaluation in individualistic cultures than in collectivistic cultures.

Socioeconomics

Our definition of culture highlights the fact that cultural differences in brand extension evaluation are likely to exist not only between national cultures, but also within a particular culture. Roth (1995) acknowledges the fact that regions within countries significantly differ from one another in terms of income, mobility, access to media, employment, and other socioeconomic characteristics. His research examines the effects of socioeconomics—in addition to cultural variables—on the performance of global brand image strategies. According to Roth (1995), a region's socioeconomic level can be assessed in terms of its level of "modernity," indicated by levels of disposable income and exposure to consumption (via TV, etc.). In markets with low levels of regional "modernity," consumers are familiar with neither the material aspects of consumption nor the ability of brands to satisfy social needs. In contrast, consumers in markets with high levels of regional "modernity" are exposed to a wide range of media and do not lack mobility.

Integrating these findings into our discussion of brand extension evaluation, it can be hypothesized that the *perceived similarity* between the parent brand and its extension is less likely to serve as an evaluative determinant for less modern regional cultures. Again, consumers in such cultures spend their income for more basic needs and hence are not as materialistic as consumers in modern regional cultures. How similar a brand extension is perceived to its parent brand, therefore, seems relatively less important to this cultural group. In contrast, consumers in modern regional cultures value symbolic and sensory consumption as a way to identify with the culture to which they have been exposed (Keyfitz 1982; Roth 1995). In this case, the fit between the parent brand and its extension should be more important because consumers from modern regional cultures are more knowledgeable and concerned about global brands in general (Alden, Steenkamp, and Batra 2006).

P24 (Perceived similarity): Perceived similarity between the parent brand and its extension constitutes a more important determinant of brand extension evaluation in modern than in less modern regional cultures.

Furthermore, Roth (1995) demonstrated that low regional socioeconomics enhances product performance when functional *brand concepts* are emphasized and that high regional socioeconomics enhances product performance when symbolic *brand concepts* are emphasized. Accordingly, consumers from cultures low in regional socioeconomics (less modern regional cultures) aim to buy products because of their functionality and performance rather than symbolic value. Again, such consumers are not familiar with the ability of brands to satisfy social needs and the aspects of material consumption. Thus, such consumers buy products rather than brands. The primary determinant of such consumers' brand extension decision-making process can be expected to be a product's functionality. It is likely, therefore, that consumers from less modern regional cultures prefer brand extensions with a functional (as opposed to symbolic) brand concept, irrespective of the level of incongruity.

In contrast, due to their wide exposure to different media and mobility, consumers from cultures high in regional socioeconomics (modern regional cultures) are materialistic and aspire to belong to certain social groups. Hence, symbolic brands constitute an important means to satisfy their needs and social belongingness. It can be expected, therefore, that consumers from modern regional cultures prefer close (congruent) brand extensions of brands with a symbolic (versus functional) brand concept because affect and knowledge transfer from the highly regarded parent brand to its extension can easily take place. Thus, a congruent brand extension fulfills—similar to its parent brand—such consumers' (those from modern regional cultures) objective to be identified with a particular social group. On the other hand, however, incongruent brand extensions of symbolic parent brands may not be easily associated with the parent brand, thereby lacking the brand extensions' ability to constitute a social sign for the belongingness to a certain aspiration group. In this case, therefore, consumers' brand extensions decision (symbolic versus functional brand concept of incongruent brand extensions) is not clear.

P25a (Brand concept consistency): In less modern regional cultures, *congruent and incongruent* extensions of brands with a functional brand concept will be evaluated more favorably compared to congruent and incongruent extensions of brands with a symbolic brand concept.

P25b (Brand concept consistency): In modern regional cultures, *congruent* extensions of brands with a symbolic brand concept will be evaluated more favorably compared to congruent extensions of brands with a functional brand concept.

Moreover, because consumers in less modern regional markets are less likely to participate in materialistic consumption and are less likely to buy brands for class affiliation and social status reasons, it can be expected that such consumers are more tolerant in terms of brand concept inconsistencies, or brand-specific association inconsistencies in general, between the parent brand and its extension.

P25c (Brand concept consistency): Perceived brand concept consistency between the parent brand and its extension constitutes a more important determinant of brand extension evaluation in modern regional cultures than in less modern regional cultures.

Similarly, the lack of mobility and of frequent exposure to media of consumers in less modern regional cultures means that this group is likely to be relatively less affected by the *quality of a brand extension's parent brand*, the *number of products affiliated with the brand extension's parent brand*, and the *social risk* associated with the brand extension's product category. Again, the opposite seems to be the case for consumers in modern regional cultures, where brands are valued for their social and sensory images and reinforce consumer identification with a particular aspiration group.

P26 (Perceived quality): Perceived quality of the parent brand constitutes a more important determinant of brand extension evaluation in modern regional cultures than in less modern regional cultures.

P27 (Number of products affiliated with a brand): The number of products affiliated with a brand constitutes a more important determinant of brand extension evaluation in modern regional cultures than in less modern regional cultures.

P28 (Social risk): Perceived social risk of the brand extension's product category constitutes a more important determinant of brand extension evaluation in modern regional cultures than less modern regional cultures.

With respect to the two parent brand characteristics—perceived quality and number of products affiliated with a brand—we further propose three-way interactions, similar to the ones proposed above for Hofstede's (2001) cultural dimensions. First, a three-way interaction between perceived quality x perceived similarity x level of regional "modernity" can be proposed. Again, we do not expect cultural differences if brand extensions are high/low on both perceived quality and the number of products affiliated with a brand. As such, we expect that modern and less modern regional culture consumers will evaluate high-fit brand extensions with high-quality parent brands favorably and low-fit brand extensions with low-quality parent brands unfavorably.

However, consumers in modern regional cultures participate in symbolic and sensory consumption to identify with a particular consumer group. Thus, the parent brand's perceived quality, as a signal of a brand's prestige and reputation, seems to be more important to such consumers than the perceived fit when evaluating new brand extensions. Therefore, we expect that consumers in modern regional cultures will evaluate low-fit brand extensions of high-quality parent brands more favorably than high-fit brand extensions of low-quality parent brands. In contrast, consumers in less modern regional cultures are likely to spend their income for basic needs. Hence, a parent brand's quality and the perceived fit are not as important in brand extension evaluations, leading to similar evaluations of such extensions.

P29 (Perceived quality x perceived similarity): Consumers in modern regional cultures will evaluate high-quality and less similar extensions more favorably than poor-quality and more similar extensions. In contrast, consumers in less modern regional cultures will evaluate such extensions similarly in terms of their favorableness.

Second, we also expect a three-way interaction between the number of products affiliated with a brand x brand concept consistency x level of regional "modernity." While consumers in less modern regional cultures favor functional brand concepts, consumers in modern regional cultures favor symbolic brand concepts. Again, consumers in modern regional cultures tend to participate in materialistic consumption. For such consumers, more affiliated products may signal the parent brand's good quality and reputation, as long as these (past) extension products are consistent with their symbolic parent brand. In this case, the new extension constitutes an appropriate choice to signal one's social status and social group belongingness. However, if their brand concept is inconsistent with their symbolic parent brand's concept, it is likely that consumers in modern (versus less modern) regional cultures will pay less attention to the number of affiliated products as an indicator of a new brand extension's symbolic value. In contrast, as mentioned above, consumers in less modern regional cultures will be more tolerant in terms of, and hence less affected by, brand concept inconsistencies than consumers in modern regional cultures.

P30 (Number of products affiliated with a brand x brand concept consistency): The number of products affiliated with a symbolic brand constitutes a more important determinant of brand extension evaluation for consumers in modern than less modern regional cultures as long as the symbolic concept consistency of these (past) extension products of the parent brand is high.

Last, with respect to our second determinant of brand extension evaluation within the broader category of brand extension's product category characteristics—*financial risk*—we expect that consumers in modern regional cultures will be less affected by the brand extension's product category's perceived financial risk than consumers in less modern regional cultures. Due to

consumers' lack of disposable income in less modern regional cultures, it is likely that the perceived financial risk of the brand extensions' product category will play a more dominant role for this cultural group. In contrast, consumers in modern regional cultures have higher disposable incomes and hence, are less likely to be affected by the financial risk of the brand extension's product category.

P31 (Financial risk): Perceived financial risk of the brand extension's product category constitutes a more important determinant of brand extension evaluation in modern regional cultures than in less modern regional cultures.

Conclusion and Managerial Implications

In this paper, we have proposed a theory-based framework designed to aid in the prediction of global brand extension success in foreign markets (Table 4.1). A comprehensive framework such as the one proposed, while existent in national brand extension research, has been lacking for international researchers and brand managers. In developing this framework, we focused on three of Hofstede's (2001) proposed cultural dimensions as well as on Roth's (1995) proposed level of "modernity." Future research should extend this framework by also examining Hofstede's (2001) other two cultural dimensions (long- versus short-term orientation and masculinity versus femininity). In addition, future research should extend our framework by investigating global brand extension evaluation factors against the background of Schwartz's (1992, 1994) proposed value system.

Our proposed framework sets the stage for future brand extension research by directing attention to key variables and relationships. In addition, our developed cross-cultural framework has implications for international brand managers with respect to relationships between the parent brand and its extension, parent brand characteristics, and brand extension's product category characteristics.

Parent Brand–Brand Extension Relationship Implications

National brand extension research has consistently found that the parent brand–brand extension relationship affects brand extension evaluation. Specifically, it has been shown that the greater the perceived similarity between the parent brand and its extension, the more favorable, and hence successful, the respective brand extension will be (Aaker and Keller 1990; Boush and Loken 1991). Furthermore, it has been demonstrated that the more consistent brand concepts are between the parent brand and its extension, the more favorably evaluated the extension will be (Broniarczyk and Alba 1994; Park, Milberg, and Lawson 1991). From a cross-cultural perspective, however, perceived similarity as well as a brand's concept consistency between the parent brand and its extension are likely to constitute more important determinants of brand extension evaluation in cultures high (versus low) in power distance, uncertainty avoidance, individualism, and socioeconomics. These cross-cultural differences in perceived similarity and brand concept consistency have the following managerial implications.

First, they suggest—all else equal—that managers will be more successful in some national markets than in others when introducing more incongruent brand extensions. For example, international marketing managers are likely to be more successful with the introduction of more incongruent brand extensions in Austria (power distance index = 11), Singapore (uncertainty avoidance index = 8), or Ecuador (individualism index = 8) than in Malaysia (power distance

Table 4.1

The Cross-Cultural Brand Extension Research Agenda

Determinant category with exemplar determinants of BE evaluation	Cultural dimensions and socioeconomics			
	Power distance (high vs. low)	Uncertainty avoidance (high vs. low)	Individualism vs. collectivism (individualism vs. collectivism)	Socioeconomics (modern vs. less modern)
Relationship between the PB and the BE:				
Perceived similarity between the PB and the BE	high > low*	high > low	indiv. > collect.	modern > less modern
Brand concept consistency between the PB and the BE	high > low	high > low	indiv. > collect.	modern > less modern
PB characteristics:				
Perceived quality of the PB	high > low	high > low	indiv. < collect.	modern > less modern
Number of product categories affiliated with a brand	high > low, high < low	high > low	indiv. < collect.	modern > less modern
Brand extension's product category characteristics:				
Social risk	high > low	high > low	indiv. < collect.	modern > less modern
Financial risk	high > low	high > low	indiv. > collect.	modern < less modern

Note: PB = parent brand; BE = brand extension.

*High > low indicates that the perceived similarity between the PB and the BE constitutes a more important determinant of BE evaluation in cultures high in power distance than in cultures low in power distance.

index = 104), Greece (uncertainty avoidance index = 112), or the United States (individualism = 91).

Second, managers need to carefully examine the target market in terms of the preferred brand concept (e.g., symbolic versus functional) and brand concept inconsistency tolerance (e.g., symbolic parent brand and extension versus symbolic parent brand but functional extension). Roth (1995) has found that a brand's concept affects its market share in different cultural markets. The same can be hypothesized to be the case for brand extensions. Moreover, managers need to pay particular attention if they decide to position their new brand extension with a brand concept that is different from the parent brand's concept. Some cultures—just as is the case for perceived similarity between the parent and its extension—are more likely to be tolerant in terms of brand concept inconsistencies (e.g., Austria, Singapore, and Ecuador) than others (e.g., Malaysia, Greece, and the United States, as per the scores noted above).

However, it is unclear how these different dimensions interact with each other. For example, a culture may be relatively low in power distance, yet relatively high in individualism, uncertainty avoidance, and socioeconomics. (e.g., Germany). Similarly, a culture may be relatively low in power distance and uncertainty avoidance, yet relatively high in individualism and socioeconomics (e.g., the United States). In such instances, it is not clear whether a far, or incongruent, brand extension strategy will be evaluated more favorably even though some of the cultural dimensions would predict so. In a similar vein, it is not clear what weight these cultural dimensions would take on, if they predicted opposing outcomes (e.g., the individualism and uncertainty avoidance dimensions predict that congruent brand extensions will be successful, while the power distance and socioeconomics dimensions predict that incongruent brand extensions will be successful). It is possible that all of these dimensions will take on equal weight in terms of their immediate effect on brand extension evaluation, or that some cultural dimensions will have relatively more weight than others. As a result, future cross-cultural brand extension research should also examine the interplay between the different cultural dimensions and their individual as well as aggregated effects on brand extension evaluation. Moreover, it seems likely that the different determinants have variable "importance weights" and that these weights are not the same from culture to culture and industry to industry. Future research, therefore, should also address this issue.

Implications of Parent Brand Characteristics

Brand extension research has furthermore demonstrated that parent brand characteristics affect brand extension evaluation. Specifically, it has been found that the perceived quality, or reputation, of the parent brand as well as the number of products affiliated with a brand positively affect brand extension evaluation and eventually success (Dacin and Smith 1994; Völckner and Sattler 2006; Smith and Park 1992). From a cross-cultural perspective, however, we have uncovered the possibility that the two determinants of perceived quality and number of products affiliated with a brand might play a significantly more important role in some national markets than in others. Accordingly, we have hypothesized that the perceived quality of the parent brand will be a more important determinant of brand extension evaluation in markets whose cultures are high (versus low) on power distance, uncertainty avoidance, collectivism, and socioeconomics. Similarly, we hypothesized that the number of products affiliated with the parent brand will constitute a more important determinant of brand extension evaluation in cultures high on uncertainty orientation and collectivism, and low on socioeconomics. It is unclear what effect power distance has on the relationship between the number of products affiliated with the parent brand and brand extension evaluation. Overall, our findings suggest the following for international marketing managers.

First, managers should assess the parent brand's perceived quality, or reputation, among target market consumers, as well as the target market consumers' perceptions about the number of products affiliated with the parent brand. Managers should then ensure that the product characteristics, which are more important in some cultures than in others, are satisfactorily perceived. For example, a high-quality parent brand in the target market is more important in cultures high in power distance (e.g., Malaysia), uncertainty avoidance (e.g., Greece), and collectivism (e.g., China or Ecuador). As a result, international marketing managers should assure that the parent brand's perceived quality, or reputation, is sufficient in a given national market for the brand extension to be favorably evaluated and hence successful. Similar reasoning can be put forward for the number of products affiliated with the parent brand.

Second, the previous reasoning implies that it might be helpful and strongly advisable for managers—depending on the culture—to support the introduction of brand extensions with a communication strategy that aims at highlighting the good quality of the parent brand and/or the several products that are already affiliated with the parent brand. This seems to be of particular importance for cultures high in uncertainty avoidance and collectivism. While some national brand extension research has examined several communication strategies (e.g., Barone, Miniard, and Romeo 2000; Bridges, Keller, and Sood 2000; Lane 2000), no brand extension study exists that examines this issue across cultures.

Again, however, it is unclear how the different cultural dimensions interact with each other and affect the relationship between the parent brand characteristics and brand extensions evaluation. In addition, it is unclear how the two determinants of perceived quality, or reputation, and number of products affiliated with the parent brand interact with each other and in combination affect brand extension evaluation (against the background of the cultural dimensions). More research is needed to understand these interaction processes.

Implications of Brand Extension's Product Category Characteristics

Prior research on brand extensions has also found that the brand extension's product category characteristics affect brand extension evaluation. In particular, it has been found that consumers' perceived social and financial risk of the brand extension category affect consumers' evaluation of brand extensions (DelVecchio and Smith 2005; Völckner and Sattler 2006). However, we have argued that this relationship is likely to be moderated by culture. As a result, we have hypothesized that the perceived social risk of the extension category constitutes a more important determinant of brand extension evaluation in cultures high in power distance, collectivism, and socioeconomics than in cultures low on these dimensions. Similarly, consumers' perception of financial risk of the extension category is likely to constitute a more important determinant of brand extension evaluation in cultures high in power distance, uncertainty avoidance, individualism, and low in socioeconomics than in cultures that are the opposite on these dimensions. This finding has the following implication for international marketing managers.

Managers need to make sure that the introduction of the brand extension goes along with an assessment of the target market consumers' perceived risk (social and financial) of the product category, into which the brand is intended to be extended. In some cultures (e.g., high power distance and uncertainty avoidance), risk perception plays an important determinant of brand extension evaluation and needs to be reduced in case it is perceived too high. In other cultures (e.g., low power distance and uncertainty avoidance), risk perception plays a less important role in the evaluation of brand extensions, and hence success, and might not need to be addressed to make consumers accept and buy the new brand extension. As a result, it is the international marketing

managers' task to assess the level of perceived risk of the brand extension product category in the target market and to decide whether marketing media need to be used to reduce the perceived risk of the extension category. Again, it is unclear how the different cultural dimensions interact and in combination affect brand extension evaluation. Therefore, further research should address this issue, too.

Overall, international managers need a tool that enables them to decide which of the determinants of brand extension evaluation are most important in a particular culture according to individual dimensions as well as a combination of these dimensions. A main objective of cross-cultural brand extension research should be to undertake the empirical analyses required to more fully develop such a conceptual tool for international marketing managers.

References

Aaker, David A. and Erik Joachimsthaker. 1999. "The Lure of Global Branding." *Harvard Business Review* 77 (November/December): 137–144.

Aaker, David A., and Kevin Lane Keller. 1990. "Consumer Evaluations of Brand Extensions." *Journal of Marketing* 54 (1): 27–41.

———. 1993. "Interpreting Cross-Cultural Replications of Brand Extension Research." *International Journal of Research in Marketing* 10 (1): 55–59.

Alden, Dana L., Jan-Benedict E.M. Steenkamp, and Rajeev Batra. 2006. "Consumer Attitudes toward Marketing Globalization: Antecedent, Consequent and Structural Factors." *International Journal of Research in Marketing* 26 (3): 227–239.

Barone, Michael J., Jean B. Miniard, and Paul W. Romeo. 2000. "The Influence of Positive Mood on Brand Extension Evaluations." *Journal of Consumer Research* 26 (March): 386–400.

Batra, Rajeev, Venkatram Ramaswamy, Dana L. Alden, Jan-Benedict E. M. Steenkamp, and S. Ramachander. 2000. "Effects of Brand Local and Nonlocal Origin on Consumer Attitudes in Developing Countries." *Journal of Consumer Psychology* 9 (2): 83–95.

Bearden, William O., and Michael J. Etzel. 1982. "Reference Group Influence on Product and Brand Purchase Decisions." *Journal of Consumer Research* 9 (September): 183–194.

Belk, Russell W. 1988. "Third World Consumer Culture." In *Marketing's Contribution to Development Research in Marketing*, ed. E. Kumcu and A.F. Firat, 103–127. Greenwich, CT: JAI Press.

Belk, Russell W., Kenneth D. Bahn, and Robert N. Mayer. 1982. "Developmental Recognition of Consumption Symbolism." *Journal of Consumer Research* 9 (1): 4–17.

Bottomley, Paul A., and John R. Doyle. 1996. "The Formation of Attitudes Towards Brand Extensions: Testing and Generalising Aaker and Keller's Model." *International Journal of Research in Marketing,* 13 (4), 365–77.

Bottomley, Paul A., and Stephen J.S. Holden. 2001. "Do We Really Know How Consumers Evaluate Brand Extensions? Empirical Generalizations Based on Secondary Analysis of Eight Studies." *Journal of Marketing Research* 38 (4): 494–500.

Boush, David M., and Barbara Loken. 1991. "A Process-Tracing Study of Brand Extension Evaluation." *Journal of Marketing Research* 28 (1): 16–28.

Boush, David M., S. Shipp, B. Loken, E. Gencturk, S. Crockett, E. Kennedy, B. Minshall, D. Misurell, L. Rochford, and J. Stroble. 1987. "Affect Generalization to Similar and Dissimilar Brand Extensions." *Psychology & Marketing* 4 (3): 225–237.

Bridges, Sheri, Kevin Lane Keller, and Sanjay Sood. 2000. "Communication Strategies for Brand Extensions: Enhancing Perceived Fit by Establishing Explanatory Links." *Journal of Advertising* 29 (4): 1–11.

Broniarczyk, Susan M., and Joseph W. Alba. 1994. "The Importance of the Brand in Brand Extension." *Journal of Marketing Research* 31 (2): 214–228.

Business Week. 2006. "The Best Global Brands." August 7, 54–56.

Chiu, Liang-Hwang. 1972. "A Cross-Cultural Comparison of Cognitive Styles in Chinese and American Children." *International Journal of Psychology* 7 (4): 235–242.

Czellar, Sandor. 2003. "Consumer Attitude Toward Brand Extensions: An Integrative Model and Research Propositions." *International Journal of Research in Marketing* 20 (1): 97–115.

Dacin, Peter A., and Daniel C. Smith. 1994. "The Effect of Brand Portfolio Characteristics on Consumer Evaluations of Brand Extensions." *Journal of Marketing Research* 31 (2): 229–242.
Dahl, Stephan. 2004. "Intercultural Research: The Current State of Knowledge." Working paper, Middlesex University, Business School London.
DelVecchio, Devon, and Daniel C. Smith. 2005. "Brand-Extension Price Premiums: The Effects of Perceived Fit and Extension Product Category Risk." *Journal of the Academy of Marketing Science* 33 (2): 184–196.
Erdem, Tülin, Joffre Swait, and Ana Valenzuela. 2006. "Brands as Signals: A Cross-Country Validation Study." *Journal of Marketing* 70 (January): 34–49.
Escalas, Jennifer Edson, and James R. Bettman. 2003. "You Are What They Eat: The Influence of Reference Groups on Consumers' Connections to Brands." *Journal of Consumer Psychology* 13 (3): 339–348.
Grewal, Dhruv, Jerry Gottlieb, and Howard Marmorstein. 1994. "The Moderating Effects of Message Framing and Source Credibility on the Price-Perceived Risk Relationship." *Journal of Consumer Research* 21 (1): 145–53.
Han, Jin K., and Bernd H. Schmitt. 1997. "Product-Category Dynamics and Corporate Identity in Brand Extensions: A Comparison of Hong Kong and U.S. Consumers." *Journal of International Marketing* 5 (1): 77–92.
Hannerz, Ulf. 1992. *Cultural Complexity: Studies in the Social Organization of Meaning.* New York: Columbia University Press.
Harrell, Gilbert D. 1986. *Consumer Behavior.* San Diego, CA: Harcourt Brace.
Hem, Leif E., Leslie de Chernatony, and Nina Iversen. 2003. "Factors Influencing Successful Brand Extensions." *Journal of Marketing Management* 19: 781–806.
Hermans, Hubert J.M., and Harry J.G. Kempen. 1998. "Moving Cultures." *American Psychologist* 53 (10): 1111–1120.
Hofstede, Geert. 1991. *Cultures and Organizations: Software of the Mind.* New York: McGraw-Hill.
———. 2001. *Culture's Consequences—Comparing Values, Behaviors, Institutions, and Organizations across Nations.* Thousand Oaks, CA: Sage.
Holt, Douglas B., John A. Quelch, and Earl L. Taylor. 2004. "How Global Brands Compete." *Harvard Business Review* 82 (9): 68–75.
Interbrand. 2006. "Best Global Brands 2006—A Ranking by Brand Value." Report. Available at www.ourfishbowl.com/images/surveys/BGB06Report_072706.pdf (accessed November 2006).
Jain, Shailendra Pratap, Kalpesh Kaushik Desai, and Huifang Mao. 2007. "The Influence of Chronic and Situational Self-Construal on Categorization." *Journal of Consumer Research* 34 (1): 66–67.
Ji, Li-Jun, Kaiping Peng, and Richard E. Nisbett. 2000. "Culture, Control and Perception of Relationships in the Environment." *Journal of Personality and Social Psychology* 78 (May): 943–955.
John, Deborah Roedder, Barbara Loken, and Christopher Joiner. 1998. "The Negative Impact of Extensions: Can Flagship Products Be Diluted?" *Journal of Marketing* 62 (January): 19–32.
Keyfitz, Nathan. 1982. "Development and the Elimination of Poverty." *Economic Development and Cultural Change* 30 (3): 649–670.
Lane, Vicki R. 2000. "The Impact of Ad Repetition and Ad Content on Consumer Perceptions of Incongruent Extensions." *Journal of Marketing* 64 (April): 80–91.
Loken, Barbara, and Deborah Roedder John. 1993. "Diluting Brand Beliefs: When Do Brand Extensions Have a Negative Impact?" *Journal of Marketing* 57 (July): 71–84.
Mandel, Naomi. 2003. "Shifting Selves and Decision Making: The Effects of Self-Construal Priming on Consumer Risk-Taking." *Journal of Consumer Research* 30 (June): 30–40.
Mandler, George. 1982. "The Structure of Value: Accounting for Taste." *Affect and Cognition: The 17th Annual Carnegie Symposium,* eds. Margaret S. Clark and Susan T. Fiske, 3–36. Hillsdale, NJ: Lawrence Erlbaum.
Merz, Michael A., and Dana L. Alden. 2007. "Uncertainty Orientation and the Evaluation of Brand Extensions." Working paper, University of Hawaii.
Merz, Michael A., Yi He, and Dana L. Alden. Forthcoming. "A Categorization Approach to Analyzing the Global Consumer Culture Debate." International Marketing Review.
Meyers-Levy, Joan, and Alice M. Tybout. 1989. "Schema Congruity as a Basis for Product Evaluation." *Journal of Consumer Research* 16 (1): 39–54.
Milberg, Sandra J., C. Whan Park, and Michael S. McCarthy. 1997. "Managing Negative Feedback Effects Associated with Brand Extensions: The Impact of Alternative Branding Strategies." *Journal of Consumer Psychology* 6 (2): 119–140.

Miller, J.G. 1997. "Theoretical Issues in Cultural Psychology." In *Handbook of Cross-Cultural Psychology*, ed. Y.H. Poortinga J.W. Berry, and J. Pandey. Vol. 1. Boston: Allyn & Bacon.

Monga, Alokparna Basu, and Deborah Roedder John. 2004. "Consumer Response to Brand Extensions: Does Culture Matter?" *Advances in Consumer Research,* 31, 216–222.

———. 2007. "Cultural Differences in Brand Extension Evaluation: The Influence of Analytic versus Holistic Thinking." *Journal of Consumer Research* 33 (March): 529–536.

Ng, Sharon and Michael Houston. 2004. "Exemplars and Beliefs? Implications of Representational Differences on Brand Evaluations Across Cultures." *Advances in Consumer Research* 31: 223.

Nisbett, Richard E., David H. Krantz, and Christopher Jepson. 1983. "The Use of Statistical Heuristics in Everyday Inductive Reasoning." *Psychological Review* 90 (4): 339–363.

Nisbett, Richard E., Kaiping Peng, Incheol Choi, and Ara Norenzayan. 2001. "Culture and Systems of Thought: Holistic Versus Analytic Cognition." *Psychological Review* 108 (April): 291–310.

Ozanne, Julie L., Merrie Brucks, and Dhruv Grewal. 1992. "A Study of Information Search Behavior during the Categorization of New Products." *Journal of Consumer Research* 18 (March): 452–463.

Park, C. Whan, Bernard J. Jaworski, and Deborah J. Maclnnis. 1986. "Strategic Brand Concept-Image Management." *Journal of Marketing* 50 (4): 135–145.

Park, C. Whan, Sandra Milberg, and Robert Lawson. 1991. "Evaluation of Brand Extensions: The Role of Product Feature Similarity and Brand Concept Consistency." *Journal of Consumer Research* 18 (2): 185–193.

Roth, Martin S. 1995. "The Effects of Culture and Socioeconomics on the Performance of Global Brand Image Strategies." *Journal of Marketing Research* 32 (2): 163–175.

Schwartz, Shalom H. 1992. "Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries." In *Advances in Experimental Social Psychology*, ed. M. Zanna, Vol. 35. San Diego, CA: Academic Press.

———. 1994. "Beyond Individualism/Collectivism: New Cultural Dimensions of Values." In *Individualism and Collectivism: Theory, Method and Applications,* ed. U. Kim, H.C. Triandis, C. Kagitcibasi, S-C. Choi, and G. Yoon, 85–119. Newbury Park, CA: Sage.

Smith, Daniel C., and C. Whan Park. 1992. "The Effects of Brand Extensions on Market Share and Advertising Efficiency." *Journal of Marketing Research* 29 (3): 296–313.

Spencer-Oatey, H. 2000. *Culturally Speaking: Managing Rapport in Talk Across Cultures.* London: Continuum.

Stayman, Douglas M., Dana L. Alden, and Karen H. Smith. 1992. "Some Effects of Schematic Processing on Consumer Expectations and Disconfirmation Judgments." *Journal of Consumer Research* 19 (2): 240–255.

Steenkamp, Jan-Benedict E.M., Rajeev Batra, and Dana L. Alden. 2003. "How Perceived Brand Globalness Creates Brand Value." *Journal of International Business Studies* 34 (1): 53–65.

Sujan, Mita, and James R. Bettman. 1989. "The Effects of Brand Positioning Strategies on Consumers' Brand and Category Perceptions: Some Insights from Schema Research." *Journal of Marketing Research* 26 (November): 454–467.

Sunde, Lorraine, and Roderick J. Brodie. 1993. "Consumer Evaluations of Brand Extensions: Further Empirical Results." *International Journal of Research in Marketing* 10 (1): 47–53.

Völckner, Franziska, and Henrik Sattler. 2006. "Drivers of Brand Extension Success." *Journal of Marketing* 70 (2): 18–34.

Wernerfelt, Birger. 1988. "Umbrella Branding as a Signal of New Product Quality: An Example of Signaling by Posting a Bond." *Rand Journal of Economics* 19 (Autumn): 458–466.

Yoon, Yeosun, and Zeynep Gurhan-Canli. 2004. "Cross-Cultural Differences in Brand Extension Evaluations: The Effect of Holistic and Analytical Processing." *Advances in Consumer Research* 31: 224.

CHAPTER 5

A REVIEW OF EYE-TRACKING RESEARCH IN MARKETING

MICHEL WEDEL AND RIK PIETERS

Abstract

Motivated from the growing importance of visual marketing in practice, we review eye-tracking research to evaluate its effectiveness. We provide a case study of the application of eye-tracking to ad pretesting. We review eye-tracking applications in advertising (print, TV, and banner), health and nutrition warnings, branding, and choice and shelf search behaviors. We then discuss findings, identify current gaps in our knowledge, and provide an outlook on future research.

Introduction

Consumers are exposed daily to hundreds of advertisements on television, in newspapers, magazines, the yellow pages, and on Internet sites, as well as to a host of other visual marketing stimuli. Companies track closely what consumers see, in order to help render their visual marketing efforts more effective. As a consequence, there has been a rapid growth in commercial applications of eye-tracking technology in the United States and Europe. Firms such as Kraft Foods, PepsiCo, Pfizer, P&G, and Unilever are leading users of the methodology. Providers of eye-tracking data now conduct hundreds of studies each year. Such commercial research companies include, for example, Perception Research Services in the United States and Verify International in Europe. The growth of eye-tracking is in large part driven by technological innovations in the development of eye-tracking devices and sharp declines in the costs of these devices. Until recently, the commercial use of eye-tracking was cumbersome, time consuming, and expensive. This situation has changed in recent years due to new generations of infrared eye-trackers, which enable eye movement recording of consumers under natural exposure conditions, with large amounts of stimuli, and at high precision and low cost. Therefore, our goal is to provide a review of the applications of eye-tracking research in marketing to date, and the insights that it has yielded. We first provide some background on eye-tracking and visual attention, then provide a case study on the use of eye movements, and proceed to review applications of eye-tracking in a number of areas of marketing.

Theoretical Background

In order to closely process a specific object or location in a visual marketing stimulus, consumers have to move their eyes. This is required because acuity across the retina of the human eye rapidly falls off with increased eccentricity from the fovea, which is the central and most sensitive part of the retina directly opposite of the lens. This makes it interesting to study eye movements as indicators

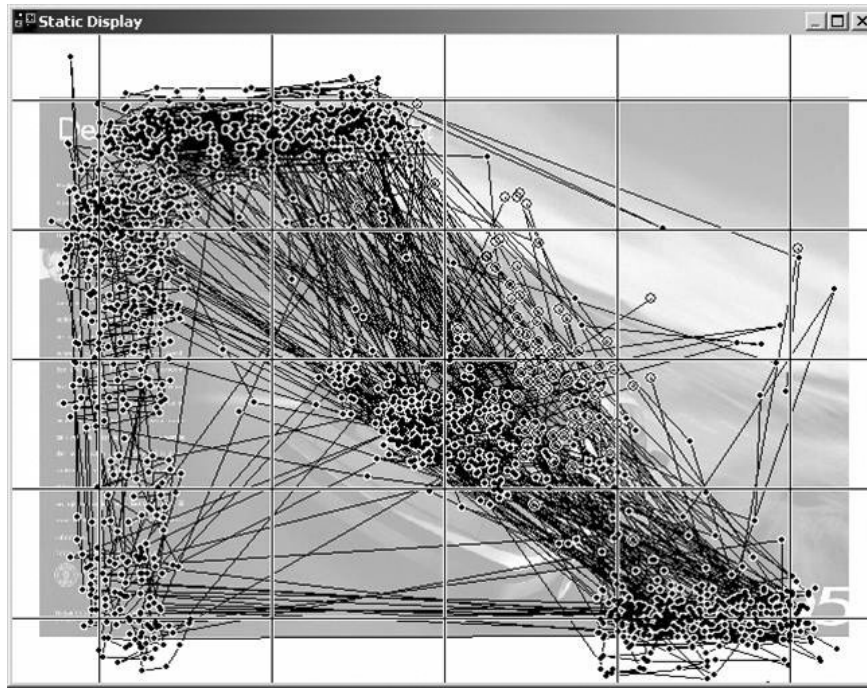
of information acquisition behavior (Russo 1978). What we believe to be smooth movements of our eyes in fact consist of two very different components: fixations and saccades (Buswell 1935). Saccades are rapid, ballistic jumps of the eyes, typically lasting around 20–40 milliseconds, that serve to project specific locations of the scene onto the fovea. The saccade is the fastest movement in the human body, and humans make around one hundred and seventy thousand of them a day. Fixations are moments during which the eye is relatively still, typically lasting around 200–500 milliseconds (Rayner 1998). During a fixation a contiguous area of the scene is projected onto the fovea for detailed visual processing. At any given point in time, only about 8 percent of the visual field is projected on the fovea and available for detailed processing.

Yet the amount of information that is transmitted through the optic nerve exceeds what the brain can process, so the brain has evolved attentional mechanisms that select a subset of relevant information for enhanced processing. When attention selects a particular location or object in a scene, processing of it is enhanced, and processing of nonselected locations and objects is simultaneously suppressed. In principle, attention may operate on spatial locations, visual features, or objects in the scene. Both traits and states of the consumer (top-down factors) and characteristics of the visual marketing stimulus (bottom-up factors) can influence attention. Much is known about the influence of stimulus-based (bottom-up) factors on attention, and that knowledge can be gainfully applied in the design of visual marketing stimuli. An object "pops out" and is found instantaneously on the first eye fixation, based on pre-attentive processes, when it stands out in the scene because of a single perceptual feature, such as the way the new Heinz green ketchup stands out among the homogeneously red competing brands on the shelf. These basic perceptual features include color, edges, luminance, shapes, and sizes of objects in a scene. The bottom-up control of attention by those features is largely involuntarily. Much less is known about the influence of top-down, voluntary mechanisms in guiding attention. Yet, both bottom-up and top-down processes influence attention. Top-down factors such as consumers' search goals ("find the new Heinz bottle," or "find the cheapest cake mix") and memory for basic features ("what is the color of the new Heinz bottle?") influence visual processing by selectively enhancing visual features that are diagnostic and by selectively suppressing features that are nondiagnostic. Such voluntary, top-down enhancement and suppression of perceptual features is effortful, usually slower. Thus, attention prioritizes objects in commercial scenes as a result of the integrated effect of bottom-up and top-down factors. Attention is reflected in eye-movements. Information extracted during fixations eventually contributes to downstream marketing effects of interest such as learning (memory), preference formation, choice, and sales.

The pattern of fixations and saccades across a stimulus such as an advertisement is called a scanpath (Noton and Stark 1971). Eye-trackers record the patterns of fixations and saccades that consumers make across a visual stimulus. The most common types for commercial purposes use infrared corneal reflection methodology, which measures the distance and angle of the reflection of infrared light from the center of the pupil to determine the point of fixation of the person, after calibration (Young and Sheena 1975). Such infrared eye-tracking has a temporal and spatial resolution of sufficient accuracy for commercial and academic applications in marketing. Modern table-mounted devices (e.g., Tobii Systems[1]) integrate miniature head- and eye-tracking cameras into an LCD monitor, which enables individuals to freely move their heads during eye-tracking. The comparatively low costs of new generations of eye-tracking systems, their short calibration times, natural exposure conditions, and unobtrusive measurement have greatly contributed to the recent growth of eye-movement applications in marketing practice as well as their use for theory development and testing in academic research.

Figure 5.1 shows an eye-tracking device (top), and the eye movements of a sample of consumers across a magazine ad (bottom). Eye-tracking provides moment-to-moment measures of the point-

Figure 5.1 **Tobii Eye-Tracker (Top), and Scanpaths of 100 Consumers Superimposed on a Car Ad (Bottom)**



of-regard of consumers exposed to visual marketing stimuli, including television commercials, print ads, catalogs, yellow pages advertising, outdoor advertising, point-of-purchase material, and Web pages. Such spatio-temporal measures of visual attention are intimately related to the attention processes of prime interest, and cannot be obtained otherwise. Because it is increasingly evident that attention is more central to the processing and effectiveness of visual stimuli than previously believed (see subsequent discussion), and because the current technology enables measurements at unprecedented scale and precision, we see a prominent role for eye-tracking in theory development and visual marketing practice. To illustrate the use of eye-tracking in visual marketing practice, we provide an application in an advertising pretest context to improve print ad layout, below.

Surprisingly, in view of the dominance of visual stimuli in marketing, the application of eye-tracking methods in practice, and, as illustrated, the importance of visual processing in consumer behavior, a perusal of the marketing literature reveals a lack of continued focus on eye-tracking. In the past, the study of visual attention in the academic marketing literature may have been hampered by the accepted wisdom that (1) attention is a mere precondition, a gate through which information enters on its way to higher-order cognitive processes of more interest, (2) gaining and retaining attention is easy, for example, through contrast with competitors and repetition, and (3) measuring attention with eye-tracking is difficult because of the costs of earlier eye-tracking generations, and unnecessary because measuring it through memory measures is easy and valid.

Academic research has shown that the study of visual attention is important in its own right and that the accepted wisdom is flawed. First, psychological research reveals that visual attention is not only a gate, as suggested by hierarchical processing models such as AIDA (Attitude-Interest-Desire-Action; Starch 1923; Strong 1920), but also and more important, a key coordinating mechanism that serves to maintain information processing and other goals over time (LaBerge 1995). Attention might even be closer to actual behavior than intuition informs us, and eye movements could be more than the tip of the iceberg (Russo 1978), as recent work by Rizzolatti and his colleagues about the potentially close correspondence of eye movements and higher-order cognitive processes reveals (Rizzolatti, Riggio, and Sheliga 1994).

Second, gaining and retaining attention is difficult, because of the high levels of advertising clutter—that is, the number of ads that consumers are exposed to at a given location or time in newspapers, magazines, or Web pages (Burke and Srull 1988; Keller 1991; Kent 1993) and the number of SKUs (Stock Keeping Units) on shelves of brick-and-mortar retailers or Web pages of online retailers (Mulvihill 2002). Breaking through the clutter and gaining consumers' attention is difficult in each of those situations. As Jones (2004, p. 146) puts it, "if consumers' attention is not caught in the first place, all findings [on message appreciation and acceptance] in 'forced exposure' research are valueless."

Third and finally, although it is likely true that standard measures of attention based on post-hoc recall are biased, measuring visual attention is easy. Biel (1993, p. 29) states, "I am troubled by the fact that (a) attention is clearly an element that determines the effectiveness of advertising, but (b) recall's failings make it unacceptable as a candidate measure of the attention dimension. That would lead me to push hard for research and development to provide a useful measure of attention. My view is that a behavioural measure of attention may be a fruitful area in which to search." Eye-tracking provides such measures. Eye movements are tightly coupled with visual attention, making them eminent indicators of the visual attention process, which is now easy to assess with modern eye-tracking equipment. Eye movements are behavioral measures of the unobservable visual attention process of prime interest. "The eyes don't lie. If you want to know what people are paying attention to, follow what they are looking at" (Davenport and Beck 2001, p. 19).

Visual attention to advertising has been shown to have systematic downstream effects on brand memory, indicating its predictive validity (Pieters, Warlop, and Wedel 2002; Wedel and Pieters 2000). Further, validity is supported by experiments that demonstrate a relation between attention to brands on shelves, measured with eye-tracking, and in-store decision making (Chandon 2002; Pieters and Warlop 1999; Russo and Leclerc 1994). Finally, a preliminary indication that attention is associated with sales was provided by Treistman and Gregg (1979), who observed that one of two ads that people looked at longer in an eye-tracking experiment also attained higher sales. With such reassuring findings on predictive validity, the current ease of eye-movement recordings, and the emergence of strong theories of visual attention, the door is open for further research on and the establishment of the academic field of visual marketing, building on and extending what has become known in recent years through eye-tracking research.

A Case Study

This case study involves performance metrics and established benchmarks to improve an underperforming print ad. It was conducted in 2000 by Verify International for a full-page print ad of Robyn, a European Unilever liquid laundry detergent brand. Data collection was done at Verify's eye-tracking facilities. The design was an "after-only with control group" design; 115 (initial ad) and 109 (improved ad) regular consumers were recruited consecutively to participate in the test, and were exposed to the initial ad and the improved ad, respectively. The ads, as they would appear in a regular magazine with an editorial counterpage, were presented on NEC 21-inch LCD monitors in full-color bitmaps with a 1,280 x 1,024 pixel resolution. Participants in the test continued to a next page by touching the lower right corner of the (touch-sensitive) screen, as when paging. Infrared corneal reflection methodology was used for eye-tracking. The equipment leaves participants free to move their heads, cameras tracking both the position of the eye and head.

The top panel of Figure 5.2 shows the initial ad, with three key regions of interest (ROI), the key ad design objects (brand, text, pictorial), and eye fixations of 115 consumers (50 percent males and 50 percent females) obtained through eye-tracking, superimposed. Table 5.1 shows the average fixations on the editorial counterpage, the ad page, and its three objects, and three key metrics derived from the ad. Also shown are benchmarks, the scores of the best 30 percent of ads from the database of 6,000 recently tested ads. The three metrics used are the Crea-Score (percentage of sample attending the ad in the first ten seconds, a measure of attention retention), Brand Contact (percentage of sample fixating the brand), and Stimulus Performance (percentage of sample fixating brand, text, and pictorial). The scores are respectively 40, 48 and 30 percent (with standard errors around 4–5 percent). It is clear that the initial ad falls short of the benchmark. These results provided input to redesign the ad in terms of the sizes and locations of the key design elements (headline, pictorial, brand, and text), without affecting its creative content. Although formal procedures can be used to design the ad based on such information, here heuristic procedures were used.

The new ad was retested. The improved ad is shown in the lower panel of Figure 5.2, again with ROI and fixation points overlaid, and Table 5.1 shows the metrics for it. The retest reveals that the improved ad, even with relatively simple and inexpensive changes in the visual layout in terms of the design elements, and virtually no change in its creative content, performs much better. The Crea-Score, Brand Contact, and Stimulus Performance metrics improve to 48, 61, and 45 percent, respectively, the latter two changes being statistically significant. But even these improved metrics fall short of the benchmark of the top 30 percent of ads. Although the retest results were already encouraging, room for further ad improvement remained, and formal optimization methods to

Figure 5.2 **Robyn Test Case Results: Initial Ad (Top) and Improved Ad (Bottom), With Same Editorial Counterpage; Regions of Interest (Rounded Boxes), and Fixation Points Superimposed**

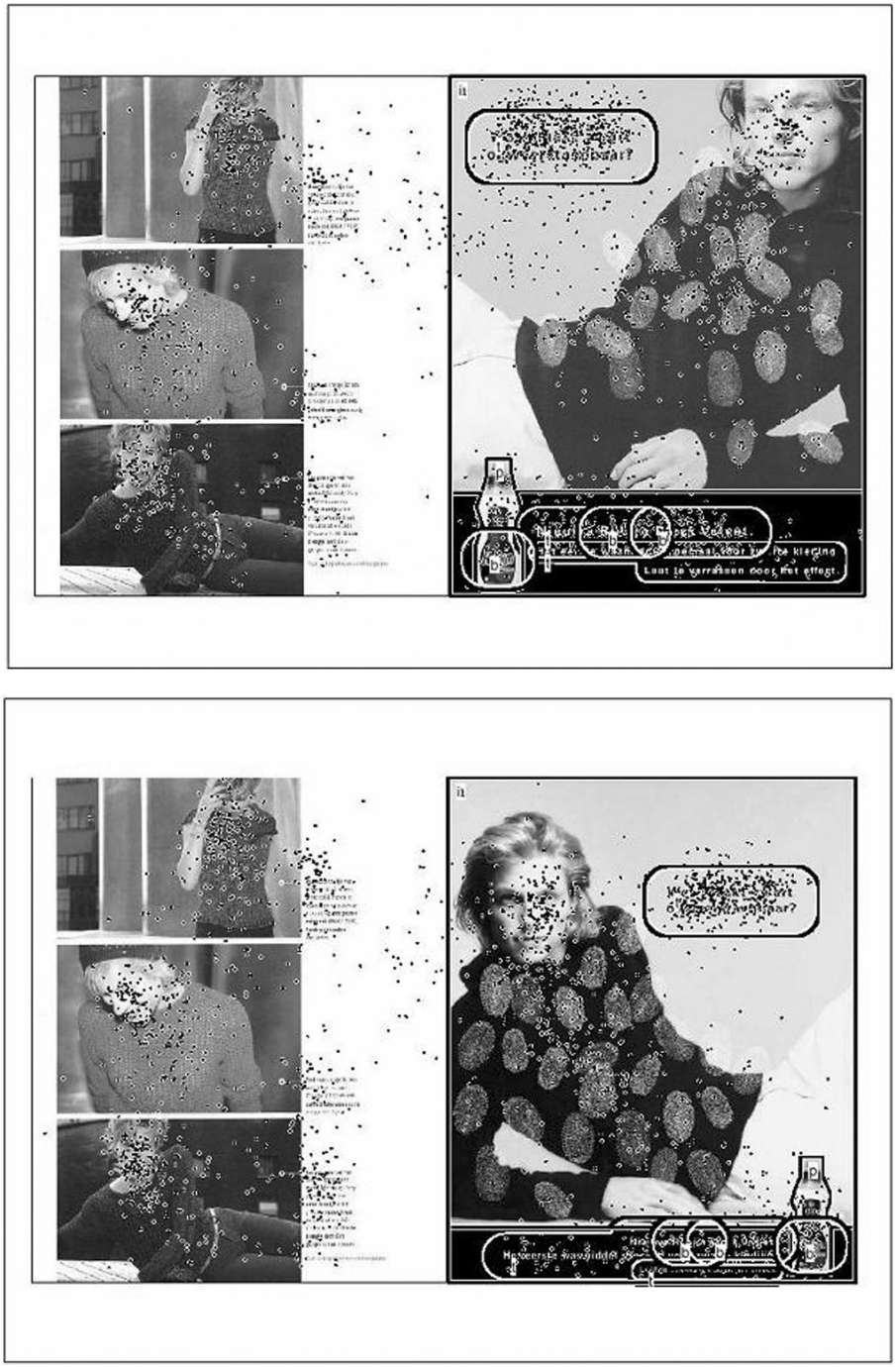


Table 5.1

Verify/Unilever Robyn Pretest Case

	Initial Ad (N = 115)	Improved Ad (N = 109)	Improvement (%)	Benchmarks
Average Fixation Frequencies				
Editorial page	7.74	6.47		
Ad page	10.63	13.22		
Brand	1.42	2.40		
Text	5.00	6.61		
Pictorial	6.83	7.19		
Performance Metrics				
Crea-Score	40	48	20	46
Brand Contact	48	61	27	83
Stimulus Performance	30	45	50	58

maximize attention by optimizing the surface sizes and locations of the key design objects in ads could be utilized. The key notion is that the eye-tracking-based attention metrics used in this case study are adequate metrics to measure ad effectiveness. The preceding discussion on the theory of visual attention and eye movements has argued this to be the case, and we will return to this issue in the discussion section. Note that the eye-tracking only takes a few seconds of participants' time to conduct, does not require verbal questioning that might disrupt ongoing ad processing, and can be done at low cost, which contributes to its increasing popularity in visual marketing practice.

Eye-Tracking Research for Visual Marketing

In 1924, in the first eye-movement analysis of print ads that we know of, Nixon observed eye movements of consumers who were paging through a magazine with print ads, while hiding himself in a box behind a curtain. Sometime later, Karslake (1940) used the Purdue Eye Camera to collect eye-movement data on advertisements appearing in the *Saturday Evening Post*. In 1950, Fitts and his colleagues (Fitts, Jones, and Milton 1950) examined the eye movements of pilots landing an airplane, which is the first usability study that provided findings now central in Web and interface research. After a period of relative silence in marketing, new impetus for the use of eye-tracking came from Russo's pioneering article of 1978, "Eye-Fixations Can Save the World," in which he argued for studying eye-movements to evaluate marketing effectiveness, focusing on consumer decision processes. Russo compared five cognitive process tracing methods, including information display boards, input-output analysis, and verbal protocols, on seven criteria. The methods were rated on seven performance attributes, including quality, validity, obtrusiveness, ease of use, and cost of the equipment. He concluded that eye-fixation methodology scores high on many of these criteria and offers advantages not offered by other methods. As process tracing data, eye movements offer detail and validity. Verbal protocols were concluded to be complementary with eye fixations, no other method being more different from eye fixations, and Russo suggested simultaneous use of these two methods in marketing research.

Since Russo's article, applications of eye-tracking have appeared in substantive areas of visual marketing involving in-store choice decisions and shelf search, print advertising, TV commercials, e-commerce, labeling and educational messages, and branding. The area of visual search alone has witnessed a steep rise from a little over ten published papers per year in the early nineties to

over two hundred a decade later. Rayner (1998) offered an extensive review of eye-tracking in reading and related areas, and Duchowski (2003) recently surveyed eye-tracking applications in various areas, including engineering and psychology, and provides a number of marketing applications as well.

In this paper, we review eye-movement research in marketing. We focus on broad categories of visual marketing stimuli and tasks: choice and search behavior, print advertising, public policy information, television commercials, Web usability and advertising, and reading tasks in survey design and branding. The fundamental differences in the nature of the stimuli in those areas allow for unique insights in the visual attention processes in question. We specifically focus on insights on the effects of bottom-up and top-down factors on the visual attention process. We examine the insights on bottom-up space-based, feature-based, and object-based attention that these studies have afforded. We also summarize the top-down role of specific goals and tasks that consumers engage in when exposed to visual marketing stimuli, for example to select one object out of multiple competing ones on a display, during search and choice, or to comprehend, evaluate, and/or memorize visual marketing stimuli as a whole, when exposed to print and television advertisements. We examine visual marketing insights gained through eye-tracking of stationary stimuli such as print ads and yellow page ads, and dynamic stimuli such as television commercials, and we explore Web usability, where consumers with multiple goals are exposed to static and dynamic scenes.

Choice and Search Behavior

In a choice or search task, the consumer's goal is to select one out of a set of multiple objects. In a choice task, preference uncertainty needs to be reduced, that is, the participant needs to decide which object to choose among the available alternatives. In a target search task, spatial uncertainty is key, that is, it is unclear where the object of interest is among its distracters. Focusing on the choice process itself, van Raaij (1977) used direct observations of eye movements applying a one-way mirror and recording camera. His study revealed a pervasive use by consumers of paired comparisons between alternatives.

Russo and Leclerc (1994) investigated the choice process for nondurables, building on earlier work by Russo and Rosen (1975). In a laboratory simulation of supermarket shelves, like van Raaij, Russo and Leclerc used direct observation of eye movements from video recordings through a one-way mirror. They heroically observed the complete scanpath of eye movements across the alternatives in the set, identifying three different stages in the choice process: orientation, evaluation, and verification, respectively. Orientation consisted of an overview of the product display. In the evaluation stage, which was the longest, direct comparisons between two or three alternative products were made. The verification stage involved further examination of the already chosen brand. Models of planned analysis of choice alternatives were disconfirmed in favor of an adaptive and constructive process. In particular, the last stage that they observed is of interest and novel, and further research may statistically test their three-state sequential model of choice.

Focusing on top-down factors, Pieters and Warlop (1999) studied the impact of time pressure and task motivation on visual attention during brand choice. All brands were new, to rule out (top-down) memory effects on choice. Analysis of eye movements revealed that visual attention adapts rapidly to differences in time pressure and task motivation, two important contextual factors, underscoring the findings of Russo and Leclerc (1994). Under high time pressure, consumers accelerated information acquisition, as revealed by a decrease in the average duration of eye fixations. Moreover, participants filtered information by skipping textual information on the packaging. Under high time pressure, consumers also shifted to a processing-by-attribute strategy,

indicated by increasing numbers of between-brand saccades. Highly motivated consumers decelerated information acquisition, indicated by longer average fixation durations, and processing-by-attribute diminished, indicated by reduced levels of inter-brand saccades. When motivation was high, consumers skipped fewer of the brand names and more of the pictorial objects. This reveals the content of consumers' brand schemas, which specify for instance that relevant information during brand choice is considered to be textual rather than pictorial. The results of this study demonstrate the potential of using eye-movement analysis to infer higher cognitive processes, the importance of task and individual factors, with implications for the pretesting of the design of packages and shelves. They also showed that brand preference can be predicted from patterns of eye movements.

In a study on the impact of point-of-purchase (POP) marketing on brand choice, Chandon (2002) develop a decision-path model of visual attention and consideration for new and existing brands. It enables estimating a product's visual (bottom-up from exposure to the display) and memory-based equity (top-down, i.e., the added value derived from these processes), and their sensitivity to person, brand, and POP marketing factors. An important finding is that looking at a brand increased its consideration probability by 30 to 120 percent. Chandon and associates concluded that the impact of visual equity is largest for brands with lower memory-based equity, which is in line with the interplay of bottom-up and top-down mechanisms in visual attention, and suggests that the current practice of allocating shelf space according to market share need not be optimal.

In four studies, Janiszewski (1998) investigated exploratory search behavior and studied relationships between the size of objects in product displays and the amount of attention devoted to them, building on and extending Anstis's (1974) theory. Exploration is similar to the orientation stage of Russo and Leclerc (1994). Janiszewski's research revealed the influence of bottom-up mechanisms in competition for attention, and demonstrated the influence of display layout, as well as of object size and eccentricity from the fixation point on attention. Part of his studies used retailer catalog displays as stimuli. He concluded that the competition for attention created by items surrounding a focal item (measured as the summed attentional demand of these distracter items) influences the amount of time a person spends looking at the item and the likelihood that the person will recall information about the item. This research has implications for visual merchandising, advertising, and catalog page layout, and shows, for example, that the current practice of increasing attention to a specific item in a catalog display by increasing its size and/or removing other items may be suboptimal, and that carefully rearranging the layout of the items could maximize attention to the display as a whole.

Lohse (1997) collected eye-movement data while thirty-two consumers chose among forty-eight businesses in a telephone directory, and recorded three choices among these listings under a hypothetical goal, such as "your car needs repair." This work is important because it synthesizes research on choice (Russo and Leclerc 1994), on brand search (Janiszewski 1998), and on attention to advertising (reviewed in the next section). A fractional factorial design was used to systematically vary color (red/black), presence of graphics (yes/no), font type (bold/plain), and ad size (small/large, display/in column), among other things, as important bottom-up factors. Analysis of the fixation data revealed that consumers scan according to serial position, but not exhaustively, and as a result, some ads are never seen. Ad size had a strong effect: consumers noticed almost all of the quarter-page display ads but only one-quarter of the plain listings. This, of course, is in line with the findings by Janiszewski (1998) on size. Participants fixated color before noncolor ads, noticed more color ads than noncolor ads, and viewed color ads 21 percent longer than equivalent ads without color. Also, they viewed 42 percent more bold listings than

plain listings. These results support the importance of basic perceptual features such as color and size, and of contrasts in these features in the design of ads and choice sets. Consumers spent 54 percent more time viewing ads for businesses that they ended up choosing, which demonstrates the importance of attention for subsequent choice behavior; consistent with the findings of Russo and Leclerc (1994) and Pieters and Warlop (1999). In a follow-up study, Lohse and Wu (2001), examined eye-movement data from Chinese consumers engaged in choosing businesses from the yellow pages of a Chinese telephone directory on a computer screen, and found convergent evidence for their previous findings, across a different culture and different ad stimuli, which underlines the generalizability of the findings. It remains the only study to date addressing cultural differences in attention and eye movements.

Print Advertising

The study of eye movements with print ads parallels studies on scene perception in psychology, which usually involves an integration of various objects in the scene into a meaningful whole (Henderson and Hollingworth 1998; Yarbus 1967). This differs from choice and search tasks where the goal is to select one out of a set of multiple objects in a display.

Kroeber-Riel (1979) cites the pioneering research in the doctoral dissertation of Witt (1977) in which sixty participants were shown advertisements on a projection screen for a fixed duration of six seconds. Of four different ads, two versions were created by manipulation of the pictorial, to invoke mild versus intense arousal. Thus, Witt examines the influence of emotions as top-down factors, and interestingly uses manipulation of the pictorial to induce these top-down effects. The average numbers of fixations on the pictorial of the two types of ads were 3.9 and 5.5 respectively, indicating more information intake under higher arousal. Fixations on the text object of the ads were not affected, however, indicating that the activation was not transferred to other objects of the ad, which Witt explained as a lack of conceptual integration of these two objects. Witt (1977) also demonstrated a relationship between fixation frequency on an object and subsequent recall of it, although the correlations were moderate (0.3–0.4).

In another early study in marketing, Treistman and Gregg (1979), both working at Perception Research Services (PRS) at the time, investigated the diagnosticity of fixation data, ability to discriminate between different print advertisements, and ability to predict which of several ads had the highest sales potential. They recorded visual selection of three ad design elements: pack shot, product copy, and price information; the proportion of time spent on each of these design elements; and the sequence in which they were visited (among other things) for a number of cosmetics ads. This study thus looks at the effects of ad objects (elements) as bottom-up factors influencing attention. The researchers observed relationships of eye-movement patterns with involvement, familiarity as a top-down influencing factor, and with purchase intent as a desirable down-stream effect. Further, combining eye-tracking data with purchase interest data correctly pinpointed the better advertisement. Although many of the findings were descriptive, this study lays an important foundation for further research in visual attention to print ads, in establishing the role of top-down factors and down-stream measures of effectiveness.

Leven (1991) conducted a study on the influence of space- and object-based aspects of the ad layout on attention. To that aim, 149 consumers were exposed to 10 print advertisements for detergent and cosmetics products under a free viewing task, while their eye movements were recorded with infrared eye-tracking. Consumers were seated about two meters from a backlit slide projector screen (1-by-1 meter) on which the ads were shown, and they could proceed to the next ad by pushing a button. Average gaze duration under these conditions was 7.2 seconds, which is

comparable to research under high involvement conditions (Rayner et al. 2001), but higher than typical gaze durations under more natural, low involvement, conditions (Pieters and Wedel 2004). Leven (1991) found that the center of ads was fixated much more frequently, with longer gaze durations than the sides, and that the upper-right corner was fixated least, and with the shortest gaze durations, which suggests that this location was only explored rather than attended to in depth. In addition, his research indicated a preferential scanpath sequence starting in the middle, going to the top, and after other steps, ending in the lower-right corner, where often the brand and/or slogan are located. Additional analyses revealed that the number of different ad objects (six were distinguished: headline, brand name, product/packshot, bodytext, people and objects) fixated rapidly increased to four after about two-and-half seconds, and than tapered off, reaching six objects after only 8 seconds. This suggests that participants quickly scanned the ads before attending to them in detail, which we return to later.

In a study on attention to print ads, Rosbergen, Pieters, and Wedel (1997) propose a methodology to account for heterogeneity in visual attention effects across consumers. For that purpose they used a mixture regression model, which accounts for unobserved segments of consumers based on their eye movements across four ad design objects—pictorial, brand, headline and body text—similar to those used by Treistman and Gregg (1979) and Lohse (1997). Thus, the primary focus here is on the bottom-up effects of size as a basic perceptual feature. In an experiment, they recorded consumers' eye-movements during natural exposure to a consumer magazine, in which experimentally designed ads were inserted. Three consumer segments were identified that exhibited distinct patterns of visual attention, respectively called the scanning (attention to headline and pictorial), initial (attention to headline, pictorial, and brand), and sustained (attention to all four objects) attention segments. Total ad viewing time increased from the first to the third segment, and the segments were shown to differ in product involvement, brand attitude, and recall, thus supporting the influence of top-down factors and the predictive validity of eye-tracking for downstream measures. By revealing how each of the three segments' attention patterns was differentially sensitive to characteristics of the ads, such as the size of their headline, this study showed the combined effects of (bottom-up) salience and (top-down) informativeness on attention to advertising. This also illustrates the importance of accounting for consumer heterogeneity in studies of visual attention, and the contribution of statistical models in making inferences on covert visual attention from overt eye-movement data.

Visual attention during repeated exposures to print advertisements was the topic of eye-tracking research on wearout by Pieters, Rosbergen, and Wedel (1999). They set out to test scanpath theory, proposed by Noton and Stark (1971), which postulates that sequences of fixations occurring upon first exposure to stationary stimuli reoccur during subsequent exposures. They proposed a statistical model comprised of submodels for attention duration and for between- and within-object saccades, with the latter described by a (heterogeneous) Markov model that captures first-order dependencies of fixations in time. They distinguish the same four ad-design objects as in Rosbergen, Pieters, and Wedel (1997) and other studies cited above. Their results showed that attention duration across advertising repetitions decreases significantly—by as much as 50 percent on average. Here the repetitions were three closely spaced exposures, mimicking the increasingly popular practice of advertising scheduling for high impact. In spite of that, and in support of scanpath theory, the switching probabilities between ad objects remained fairly constant across repetitions. These results support similar findings by Harris (1993) on the stability of scanpaths. Whereas scanpath theory specifies that the stability of scanpaths is due to storage of such paths in (visuo-spatial) memory and thus a top-down effect, it now seems more likely that the stability of scanpaths is in fact driven bottom-up by the layout of the ad, although there have been no direct tests of this speculation

yet. This research attests to the importance of visual attention in understanding mechanisms of advertising effectiveness and ad wearout.

Following up on the previously described study, Liechty, Pieters, and Wedel (2003), develop a hidden Markov model to capture the temporal dimension of visual attention to print ads. Their model is based on evidence, cited earlier, that while visually exploring print ads, consumers' attention switches between two unobserved states. The idea is that during scene perception, people try to reduce two types of uncertainty, namely, *what* the identity of the various objects in the scene is and how these contribute to the overall meaning of the scene, which requires detailed, focal attention, and *where* the various informative objects and locations in the scene are. Liechty, Pieters, and Wedel (2003) infer these two visual attention states, their relative prevalence and sequence in time, from the complete sequence of eye-tracking data. For that purpose, they develop an extension of existing hidden Markov models (formulated and estimated in the Bayesian framework). The model describes the observed time-series of saccades on a spatial (48-cell) grid overlaid on the ads, for which a first-order Markov model is used, similar to Pieters, Rosbergen, and Wedel (1999). But their account of the visual attention process is primarily spatial, as it is in Leven (1991). A hidden two-state Markov process, indicative of attention switching between global and local states, is assumed to drive those observed eye movements. Individuals are assumed to make shorter saccades in the local than in the global state. Note that this model formalizes and allows explicit tests of, among others, Russo and Leclerc's (1994) and Leven's (1991) findings of unobserved attention states sequentially driving eye movements. The model was estimated on eye-movement data collected in a study of sixty-nine consumers exposed to seventeen print advertisements in their natural context. Participants nearly always started in the local attention state (with a probability of over 90 percent), but the probability of terminating exposure to the advertisements is much larger when in the global attention state (23 percent). During exposure to an advertisement, participants jump between the local and global states 2.6 times on average and tend to spend longer in the local (1.13 seconds) than in the global (0.22 seconds) state. By switching back and forth between local and global attention, the problem of interpreting the complex visual scene is broken down into a sequence of simpler localized interpretations of the most salient and informative regions. Although the link between the two states of the model and the "where" and "what" pathways in the visual brain is not uncontested (Reichle and Nelson 2003; also see Wedel, Pieters, and Liechty 2003), taken together, these results show that the local attention state is dominant, with a high probability of starting in that state, where there are occasional short jumps to the global attention state with a high probability of ending attention to the stimulus in that state. The authors did reveal that the two-state model provides a much better description of the data than a one-state model, but did not test for a three-state model. This research demonstrates how one can use eye-movement measures as indicators of underlying top-down factors to test theories of attention, which Feng (2003) refers to as a desirable approach to "reverse inference." Such a reverse inference approach is important because it allows us to infer cognitive processes from observed eye-movements, and enables tests of attention theories proposed in psychology and cognitive neuroscience. These are facilitated by the large sample sizes of stimuli and participants, and the temporal and spatial resolution of current data collected through eye-tracking.

Among the pioneers on eye-tracking research in reading, Rayner and coauthors (2001) used print advertisements in a study on attention to text and pictorial information under various levels of involvement. Little research in psychology had addressed that issue previously. In their experiment, half of the participants were instructed to pay special attention to car ads, and the other half were told to pay special attention to skin-care ads, whereas both groups were exposed to both the car and skin-care ads. Thus, Rayner et al. chose to study top-down effects by manipulating task

instructions. Viewers tended to spend more time looking at the text (about 70 percent) than the picture part of the ad, and spent more time looking at the type of ad to which they were instructed to pay attention. Fixation durations and saccade lengths were both longer on the pictorial than the text, but more fixations were made on the text. Fixations on the text seemed to follow predictable patterns found in other studies on reading. Here, participants tended to read the large print (in the headlines) first, then the smaller print (in the body text), and then attended to the picture. Although some of the results of the study may be caused by demand effects, the participants being specifically instructed to pay attention to the ads (Yarbus 1967), a valuable finding is that the attention to advertisements is strongly influenced by the task instructions (looking at car versus skin-care ads). One of the conclusions of the study was that perception of ads and other complex scenes is largely based on textual information, and the lack of transfer from pictorial to text that was found is consistent with findings from other studies (Witt 1977). A more extensive study on top down effects on attention to print ads was reported on by Pieters and Wedel (2007). In the study, five processing goals directed at the ad as a whole or the brand element, and inducing on memory or learning were experimentally varied. The results demonstrate a marked effect of these goals on consumer eye-movements, in line with what Yarbus (1967) already reported. For example, text receives most attention under a brand learning goal, while an ad learning goal uniformly promotes attention to most elements. The headline appeared to be invariant to these top-down effects, however and received equal attention under all conditions.

Wedel and Pieters (2000) exposed 88 consumers to 65 print ads appearing in their natural context in 2 magazines, in an attempt to formalize the attention-to-memory relationship observed in several previous studies (Lohse 1997; Rosbergen Pieters, and Wedel 1997; Witt 1977). The frequency of eye fixations on the brand, pictorial, and text objects of each ad was recorded for each consumer, and consumers were later asked to identify the brands from pixilated images of the brands, in a perceptual memory task. This indirect memory task is importantly different from the less sensitive recall tasks used in previous research. Accuracy and latency of memory were assessed. A contribution of this study was that the authors provide an account of the processing that takes place to store information in long-term memory, which had been lacking in previous research. They developed a hierarchical Bayesian model for that purpose, grounded in attention and memory theories. The model describes the process by which eye fixations on print advertisements lead to memory for the advertised brands. At each fixation an information chunk is extracted from the ad, the amount of information contained in it varying randomly across ads and consumers, and is accumulated in long-term memory across multiple fixations to the ad objects. The total amount of accumulated information influences both the accuracy and latency of brand memory, and accurate memory is assumed to occur when the accumulated information exceeds a threshold. The study looks at size of ad elements as basic features that affect attention bottom-up. Wedel and Pieters confirmed again that across the two magazines, brand surface size had a particularly prominent effect on fixations. On average, the brand was much smaller than the pictorial (about 10 times) and the text (about 3 to 5 times), but the brand received by far the most fixations per unit of its surface, followed by the text object—supporting the Rayner et al. (2001) findings—and the pictorial. However, whereas fixations to the pictorial and the brand systematically promoted accurate brand memory, text fixations did not. This finding adds insight to previous results on the attention-memory relationship, but casts doubts on the assumed central role of text in advertising. The authors also found a recency and (smaller) primacy effect on memory, while controlling for attention—indicating the role of memory retrieval cues.

Pieters, Warlop, and Wedel (2002) extended the study of Wedel and Pieters by first investigating the influence of ad originality and familiarity on consumers' eye fixations on the brand, text, and picto-

rial objects of advertisements, and how these factors, along with the information extracted during eye fixations, promote memory for the advertised brand. Ad originality is used commonly in advertising to break through competitive clutter. However, it could have detrimental effects when consumers pay more attention to the ad at the expense of the advertised brand itself. Moreover, the positive effects of originality can quickly fade when the ad becomes familiar and the novelty wears off. In Pieters, Warlop, and Wedel's (2002) study, infrared eye-tracking was applied to collect eye-fixation data from 119 consumers who paged through two general-audience magazines containing 58 full-page advertisements. In support of the researchers' hypothesis, original advertisements drew more rather than less attention to the advertised brand. More important, however, advertisements that were both original and familiar attracted the largest amount of attention to the advertised brand, which improved subsequent brand memory as well. In addition, originality and familiarity were found to also promote brand memory directly. These results reveal the double benefits of ad originality and how eye-tracking research can detect this in pre- and post-testing, and thus support ad development and media planning

Whereas many of the previously described studies documented the effects of surface sizes of ad objects on visual attention, these results had limited generalizability due to a relatively small and select set of advertisements. Therefore, Pieters and Wedel (2004) conducted an analysis of 1,363 print advertisements tested with infrared eye-tracking methodology among over 3,600 consumers. Data were aggregated across consumers. They examined attention capture and transfer for three ad objects—brand, pictorial, and text. Attention capture was operationalized as the percentage of participants fixating a selected ad object at least once. Attention transfer was operationalized as the effect of gaze duration for one of the ad objects on gaze for the other objects, comparable to Janiszewski's (1998) operationalization of attention competition on displays. Each of the three ad objects had a unique effect on attention to advertisements, which are at odds with commonly held ideas in marketing practice. The pictorial is superior in capturing attention, independent of its size, which goes against recommendations in ad practice to make the pictorial size large or even as large as possible. The text object best captures attention in direct proportion to its surface size, that is, attention to the text increases most by increasing its size, among these objects. The brand object most effectively transfers attention to the other objects. That is, longer gazes to the brand object carry over to those on other objects, but not necessarily the other way around. The authors found little or no transfer from the pictorial to brand and text objects, which generalized previous observations by, among others Witt (1977) and Rayner et al. (2001), but are at odds with common thinking about the influence of the pictorial in attention guidance. Only increments in the text object's surface size produced a net gain in attention to the ad as a whole. Brand familiarity was shown to reduce attention to the brand object but simultaneously increased attention to the text object, rather than having a uniform effect across all objects. The study has obvious implications for the role of the design objects in print ad design in practice. Finally, Pieters, Wedel and Zhang (2007) investigated the optimal design of feature advertisements. These ads are particularly prone to clutter, because multiple feature ads often appear together on a single page. The authors proposed two measures of clutter, and estimated a model that allowed them to optimally design a feature ad display page. Using data from close to 1,500 newspaper feature ads, they showed that the design can be significantly improved to attract more attention to the entire ad display, by decreasing the size of the pictorial, and increasing the sizes of promotion and price elements in particular.

Public Policy and Social Marketing Information

Whereas print ads are scenes containing multimodal information for commercial purposes, common health warnings and nutrition labels are basic iconic and textual components of ads and packages.

There is a history in social marketing and public policy research of eye-tracking mandatory health warnings and nutrition labels. In a series of studies, a group of collaborators (Fischer et al. 1989; Fox et al. 1998; Fletcher et al. 1995; Krugman et al. 1994), extensively documented adolescents viewing tobacco and beer ads. These studies are interested in attention to these labels as whole objects, on which bottom-up and top-down factors may jointly operate. In one study, (Fischer et al. 1989) with sixty-one adolescents exposed to five different tobacco advertisements, the average viewing time of the warning amounted to only 8 percent of the total advertisement viewing time. Almost half of the participants did not fixate on the warning in the ad at all, while about 36 percent of the participants appeared to have read some of the warning. Following ad viewing, participants were asked to identify the warnings they were exposed to in a list that included other simulated warnings. Participants did only slightly better than random guessing in this recognition test. The authors concluded that the federally mandated warnings must be viewed as a largely ineffective public health message.

In a study by Krugman and associates (1994) among 326 adolescents, new warnings were developed and tested. Two mandated warnings and two newly developed warnings were embedded in magazine ads for two cigarette brands. Participants viewed each ad as long as desired while eye movements were recorded. Participants subsequently completed a masked recall task. The metrics used were the number of participants who noticed the warning, their time to first fixation on the warning, and the time spent fixating on the warning. Results indicated that within the competitive reading environment created by a cigarette advertisement, new warnings attract greater readership, with quicker attention to the new than to familiar, mandated warnings. The difference in attention capture was about 10 percent, the new warnings capturing more of the participants' attention; the proposed warnings also captured attention faster, times to the first fixation on the warning being between 0.8 and 2.4 seconds. Attention engagement was also longer for the new warnings, by about half a second, or 1–2 fixations. On average, participants spent 2–3 seconds viewing the warnings. The authors found the attention measures to be significantly related to a subsequent masked recall measure. Next, Fletcher et al. (1995) provided stronger evidence of the association between dwell-time and content recall, by including print ads for two nontobacco products that also contained warnings, along with the above cigarette ads and their warnings. These findings are in line with findings on attention effects on memory for print ads, in particular those by Wedel and Pieters (2000). Implications of these studies were that the effectiveness of warnings can be improved through simplicity and novelty. Simplicity can be achieved by reducing the number of words in the text (24 for the existing warnings), and novelty by a more rapid change of warning content and formats or by originality, to prevent wearout (see Pieters, Warlop, and Wedel 2002). These studies thus demonstrate the effect of both basic features and semantic content.

Fox et al. (1998) followed up on these studies and exposed 148 fourteen- to eighteen-year-olds to five frequently run print ads targeted to this segment, while recording their eye movements. Two of the ads were for cigarettes, and one was for a beer brand. Interestingly, there were significant differences in the gaze durations for the two cigarette ads that carried over to the warnings: warnings in ads that are viewed longer also receive a longer gaze (the gazes on the warnings were relatively long, 2 and 2.5 seconds, respectively). This again points to the importance of investigating the warning in the context of the ad, and to the effects of the context on attention in general (see Janiszewski 1998). The alcohol warning received less attention (1.6 seconds) than the cigarette warnings, and less than the logo in the ads.

Two studies of product labeling used visual search tasks while tracking eye movements, investigating the effect of perceptual features on search performance. Laughery et al. (1993) studied warnings on alcoholic beverage containers with the aim of providing recommendations for format specifications. In one of three experiments, the authors tracked eye movements for

twenty-four participants exposed to thirty-eight labels, selected from a factorial design with text color, pictorial color, presence of border, and an icon as factors. Thus, this study looked at the effects of basic perceptual features, bottom-up. Participants were asked to identify the warning, and response latency times were collected in a search task. The study suggests, although overall effects were small, that perceptual features used in conjunction may help improve the speed of visually detecting the warning (by 50 percent to about .5 seconds) as well as shorten the latency of responding to it.

Goldberg, Probart, and Zak (1999) evaluated food labels' ability to promote a fast and accurate visual search for nutrition information. They looked at experience as a top-down moderating factor. Participants (five practiced label readers and five nonreaders) viewed 180 nutrition labels on a computer and were asked to find answers to specific questions (e.g., about serving size) under different fixed exposure durations (1, 3, and 5 seconds) in a visual search task. Label manipulations included several alternative line arrangements, location of the target item, and label size. Dependent measures included search time and fixation frequency, as well as the accuracy and duration of the identifying fixation. Practiced label readers acquired the target more quickly and accurately than did less-practiced readers, revealing a strong top-down effect. Surprisingly, label size did not have a significant effect. Thinner as compared to thicker anchoring lines had a larger effect on visual search time. Targets near the center of the label required one-third more time and were harder to find than targets at the top or bottom of the label, which may be caused by more dense information near the center. This study, counter to the Laughery et al. (1993) findings, shows that low-level perceptual features as well as spatial characteristics may influence visual search performance.

Television Commercials

Eye-movement data across dynamic scenes such as television commercials are, what might be called, "doubly dynamic," since the eyes move over images that themselves move too. Collecting and analyzing such attention data is challenging, which may be one of the reasons for the comparative paucity of research in this area. However, a few studies did take up the gauntlet. Using televised soft drink ads as stimuli, Janiszewski and Warlop (1993) found evidence in three experiments that conditioning procedures, aimed to establish associative learning about the target brand, enhance attention to the brand. The account of visual attention is thus primarily object (brand) based, and the learning effects investigated top-down effects. This is thus an important demonstration of a top-down influence on attention in the context of dynamic stimuli. Janiszewski and Warlop concluded that it seems possible to use the sequence of scenes in a TV commercial to encourage attention to the brand through conditioning, and to simultaneously make semantic information available for comparing choice alternatives, as input into choice behavior—a speculation that future research might test.

In a study of televised advertising images during soccer games, d'Ydewalle and Tamsin (1993), following up on an earlier study by d'Ydewalle et al. (1988) used eye-tracking to investigate the incidental processing of (static) advertising billboards placed around the soccer field, and their subsequent recall and recognition. Both studies found that participants do not look at the advertising panels substantially, and because they do not look at them, they are unable to recall and to recognize them. They did find specific effects of scenes that contained close-ups of the panels, pointing to the potential of scene-based analyses, but did not investigate the effect of basic features or top-down factors. Yet these are the only academic studies that have investigated the effectiveness of billboard product placements in sport games, and should serve as an important base for future research in this area. In a second study on dynamic stimuli, d'Ydewalle, Desmet, and Van Rensbergen (1998),

investigated the effects of successive shots in movies, and the extent to which shot transitions are perceptually disruptive. Their findings are of obvious relevance to TV commercials as well. This is one of the few studies to date that investigates the bottom-up effect of movements on attention. The researchers distinguish three types of transitions between shots. First-order transitions refer either to small displacements of the camera position or to small changes of the image size, possibly disturbing the perception of apparent movement and leading to the impression of jumping. Second-order transitions follow from a reversal of the camera position, leading to a change of the left–right position of the main actors and a complete change of the background. With third-order transitions, the linear sequence of actions in the narrative story is not obeyed, and these are meant to reinforce the narrative continuity of the story. The experiment of d'Ydewalle, Desmet, and Van Rensbergen (1998) involved eye-tracking on 76 participants watching 4 experimentally designed versions of a 7-minute movie segment. As a dependent measure, d'Ydewalle and coauthors used the standard deviations of the x-y locations of the individual fixations in 200-millisecond-long segments, as a measure of attentional homogeneity or focus, caused by bottom-up effects. The results showed that there was an increased spatial variance of eye movements 200–400 milliseconds after both second- and third-order shot transitions. Such an increase was not obtained after a first-order transition, suggesting that the increased spread of the distribution of eye movements after second- and third-order transitions is due to postperceptual, cognitive effects and that these transitions disrupt perception and cognition.

Aoki and Itoh (2000) proposed a methodology for analyzing viewers' attention to television commercials. Their results, based on a small sample of four participants, suggest that product preference is not a major factor for attention to TV advertisements, but that it may be influenced more by bottom-up design factors such as scene-change frequency and the use of celebrities. In a second study, Aoki and Itoh (2001) investigate influences of auditory information on human visual attention during viewing of television commercials. They characterize patterns of gaze toward scenes with and without sounds and found that participants' gaze patterns were affected by auditory information. Clearly these initial studies provide some valuable insights, but the detailed study of attention to TV commercials remains a wide open area.

Web Usability and Advertising

E-marketing through Web stores, electronic auctions, brand sites, banner and pop-up ads on Web sites, search engine marketing, and so forth is large and growing, and has become an integral part of firm strategy and tactics. Given the growth of expenditures on e-marketing, surprisingly little academic marketing research to date has addressed this field. There is an emerging stream of research in applied psychology and engineering on human–computer interaction and Web usability (e.g., Cowen 2001; Jacob and Karn 2003), but its theories and findings have not yet been fully infused in visual marketing, and there are only a few studies on Web advertising. Because Web design combines static and dynamic visual features, and requires consumers to engage in search, choice, and scene perception tasks often simultaneously, it could become a fertile ground for testing and applying integrative theories of the human perception–action cycle and its implications for marketing.

The Poynter Institute[2] has conducted probably the best-known Web usability studies using eye-tracking. In their Eyetrack III, for instance, eye movements were recorded of the forty-six people who were examining mock news Web sites and real multimedia content. One of the findings was a spatial effect on eye movements: the eyes most often fixated first in the upper left of the page. Only after perusing the top portion of the page for some time did the eyes explore

further down the page. Dominant headlines were found to most often draw the eye first upon entering the page, which is consistent with the work of Rayner et al. (2001) in the context of print ads. However, (even) headlines had less than a second of a site visitor's attention, which underlines the speed of consumer processing and the importance of immediate communication by the largest text.

Dreze and Hussherr (2003), in a study on the design of a French Web portal among fifty participants, examined why banner ads seem to be ineffective, and what advertisers can do to improve their effectiveness. Click-through rates are still the dominant measure of Internet advertising effectiveness, but are disappointingly low in many cases, and it is thus important to understand their determinants. Dreze and Hussherr used eye-tracking to investigate Web surfers' attention to online advertising, combined with a survey of Internet users' recall, recognition, and awareness of banner advertising. They claim that their study suggests that the reason why click-through rates are low is that surfers avoid looking at banner ads during their online activities. Out of the eight banner ads that participants were exposed to, they looked at four on average. The researchers find strong location and size effects of regions of interest on the portal's Web pages, reconfirming findings on bottom-up effects of space and size from other application areas of eye-tracking research. But, they also find a negative effect of regions of interest that contain the banner ad. This may imply that part of a surfer's processing of banners will be done peripherally and/or pre-attentively. If peripheral processing does take place, the clickthrough rate may not be the optimal, or only, preferred measure to assess banner ad performance. Alternatively, these findings could mean that relative to page content, banner ads attract less attention, for example, because they are smaller or because their locations, which are stored in consumers' spatial memory, can easily be ignored without active avoidance taking place. Interestingly, "experts" spent less time looking at the Web pages and inspected fewer regions than novices, whereas their scanpaths seemed more regular, confirming the top-down effect of expertise previously documented by Goldberg, Probart, and Zak (1999) for food labels. Thus, experts appear to be more efficient and systematic in their visual exploration (see also the results of Noton and Stark [1971] on exploration of paintings by experts and novices). Men and woman behaved similarly, but older people spent more time inspecting Web pages than young ones, presumably because they have less automatized visual routines for Web handling and thus need more sustained attention. Banner ads did have an impact on traditional memory-based measures, advertising recall, brand recognition, and brand awareness. Although brand awareness is an important ad effectiveness measure, given the problems with such measures reported previously, the conclusion that these are preferred measures of ad effectiveness would need further study. Advertisers should rely more on visual attention measures rather than clickthrough (or in addition to it), which is one key conclusion to be drawn from this study.

In a test of scanpath theory (Noton and Stark 1971) for attention to repeated Internet images, Josephson and Holmes (2002) exposed eight participants repeatedly to a portal page, an advertising page, and a news story page on the Web for a duration of ten seconds each. These authors used string-editing techniques to assess the complete sequence of eye fixations across predefined areas of interest on the Web pages, and how these sequences changed across exposures. The results indicate both substantial differences between participants in their scanpath for the same Web page, which may be indicative of top-down effects, as well as substantial differences between Web pages, which may be indicative of bottom-up effects due to the Web page layout. The findings call for further research that allows for statistical tests of differences between (larger samples of) stimuli and participants to establish the contribution of top-down and bottom-up effects on attention to Web pages.

Reading Tasks in Survey Design and Brand Extension Research

Eye-tracking in Web design and Web advertising serves to improve human–computer interaction by adapting interfaces to natural human perception and decision processes. Eye-tracking can also serve such ergonomic functions in other areas, such as survey and questionnaire design in marketing research and in brand extensions, for which reading tasks have been used. In particular, for large-scale paper-and-pencil marketing and Web survey research, extensive pretesting of the main questionnaires can be conducted to detect potential problems with question comprehensibility, response formats, and questionnaire routing (Groves et al. 2004). Redline and Lankford (2001), for instance, examined how participants attend to and make use of branching instructions (e.g., to skip a number of questions), and thus the navigational paths in questionnaires. Not following the proper paths could increase item non-response or cause inaccurate responses, and could also increase irritation and confusion, all of which is costly. They offered three versions of a questionnaire about life-style issues, using different branching instructions, to twenty-five participants in a three-group design. Eye movements were recorded on the question, answer categories, check box, and branching instructions, and return sweeps between question components were examined. The analyses revealed that errors-of-commission, where branching instructions are not followed as evidenced from eye movements, varied significantly across the three conditions. One conclusion was that many branching instructions were not read at all, and that to be effective, branching instructions needed to immediately precede or follow the specific response category to which they related. This is in line with earlier work in reading research showing that most processing takes place during or immediately after eye fixations on the relevant message components ("processing immediacy"), and that thus the time lag between the instruction in the message and the required action of the consumer should be minimized.

Another emerging area where eye-movement analysis provides new insights is brand extension research. As a case in point, Stewart, Pickering, and Sturt (2004) propose an ingenious eye-movement-based technique to assess the effectiveness of brand extensions. Brand extensions are common, as much as 95 percent of all new product introductions involve them, yet failures are also common and harmful, cannibalizing existing brands and damaging brand equity. Whereas a range of established explicit methods exists for gauging the success of an extension, it is not sufficiently clear to what extent these methods can tap the fast cognitive processing during brand extension evaluation in the marketplace. The authors propose a new method based on the notion that ambiguities created by implausible extensions cause processing difficulties that should surface in eye-movement patterns. Eighteen participants were presented with three types of sentences (e.g., *I wanted to take a picture on Polaroid, but it cost too much*; *I wanted to record a song on Polaroid, but it cost too much*; *I wanted to serve a trifle on Polaroid, but it cost too much*) that could only be fully understood if they made an inference that was based on extending the meaning associated with the brand name. The three conditions were standard brand usage, the plausible brand extension, and the implausible brand extension. Stewart and coauthors predict that plausible brand extensions will cause little difficulty during initial reading, but that implausible extensions will cause immediate disruption. Eye movements of eighteen participants who were reading these sentences were tracked. The sentences were classified into brand region, spillover region (immediately following the brand), and end-of-sentence region (the remainder of the sentence). The measures used were, among others, regression-path time (the gaze from the time the region is first entered to the time it is last exited), probability of first-pass regression (proportion of times in which the region is exited to the left after a first-pass fixation in the region), and total gaze time on the region.

Analyses of these measures confirmed the predictions and yielded two main results. First, there was

evidence of difficulty in processing the implausible condition in comparison with the normal usage condition, as reflected in 200 milliseconds longer viewing times on average, which is substantial. Second, the results showed that even plausible extensions led to minor reading difficulties, in particular later during the processing of the sentence. The authors interpret this to result from the construction of a new interpretation associated with the brand. Implausible extensions, in contrast, cause immediate disruption, resulting from the difficulty of providing a coherent interpretation. The study shows that eye-tracking, due to its high temporal resolution, reveals the ease with which participants can interpret different brand extensions. The analysis, in which measurements and metrics are derived from substantive theory on the underlying cognitive process, provides a compelling demonstration of the extent to which consumers find particular brand extensions plausible.

Conclusion

This review demonstrates that since the first eye-tracking studies in marketing by Nixon (1924), a sizable body of literature has accumulated using eye-tracking technology to assess visual marketing effectiveness. With regard to bottom-up factors, that body of literature has demonstrated the effects of spatial location of visual marketing stimuli, the effect of basic features such as color, lines/edges, movement, and in particular, size, of specific objects, and of context. Top-down effects that have been demonstrated include those of memory, involvement, attitudes, processing states, emotions, goals, and expertise. And, these attentional effects have been demonstrated for exploration, search, and choice tasks. But the body of evidence on these factors is far from comprehensive or conclusive.

The research does reveal the power of eye-tracking to the development and growth of the visual marketing discipline as a unified methodology to track visual marketing effectiveness in many areas of marketing. First, eye-tracking provides insights into communication processing and effectiveness that cannot be obtained by traditional measures because of the speed and lack of conscious access to the rapid attentional processes taking place during communication exposure. For instance, the opportunities to unintrusively monitor consumers' visual attention to visual marketing stimuli from moment to moment in time revealed the influence of information intake during eye fixations on brand consideration and preference during stimulus-based choice, and the switching between attention states during ad processing. Second, eye-tracking complements traditional measures by providing more detail about the spatio-temporal dynamics of attention and it thereby enables developing and testing integrated models of visual communication processing and effectiveness. For example, combining eye-tracking measures with traditional verbal measures of brand memory allowed structural models of the attention storage and retrieval process, and tests of the influence of attention enhancement and inhibition in Web advertising. Third, eye-tracking research challenges received knowledge about the determinants of visual communication performance that have been obtained by traditional measures of communication processing and effectiveness. For example, such research has shown that current ideas about the influence of the content and size of ad elements on attention capture and transfer, and about the order of attending to ad objects, need updating.

Yet, several areas remain under-researched, and we identify five important ones here. First and foremost, more work is needed that applies eye-tracking measures to make reverse inferences about fundamental communication processes. To date, eye-tracking research has often been descriptive, for instance, by relating perceptual aspects of the communication stimulus, such as the size, color, or position of the brand, text, or pictorial, directly to measures of visual attention. While this has led to many new insights, much more can be gained from using eye-tracking measures as indicators

of the latent processes of interest, taking a structural-theory-driven approach to modeling them. Such a reverse inference approach (Feng 2003), in which key indicators of unobservable cognitive processes are derived from the observed eye movements, will not only likely lead to more insights into visual marketing effectiveness, but will also enable realistic tests of attention theories from cognitive neuroscience and perception psychology that so far have had to rely on abstract stimuli and impoverished exposure conditions. We believe that integrating eye-tracking data in a *reverse inference* approach holds much promise for the theory and practice of visual marketing.

Second, more research about the influence of stimulus informativeness on visual attention is called for. To date, eye-tracking research in marketing has typically examined the influence of salience, as a stimulus-characteristic, residing in local contrasts in perceptual features such as size, luminance, color. As a consequence, much less is known about the role that the informativeness of objects to the consumers' goals play in attention capture and retention. Carefully designed experiments and statistical models need to be combined to disentangle effects of attention to features, space, and objects in commercial scenes from observed eye movements. Evidently, top-down factors such as consumer traits and states have an impact on attention patterns, and some states may even be primed bottom-up (endogenously) by certain objects in the visual marketing stimuli themselves, such as when attention to a snack brand in a retailer's billboard primes a specific weight-loss goal in consumers, or when the text in print ads primes a more general brand-learning goal. The pioneering work of Alfred Yarbus (1967) on the influence of goals on the attention patterns is a landmark that can be build upon. The final attention patterns of consumers across visual marketing stimuli are the product of the salience and the informativeness of objects to the consumer, and we know very little about the latter. More *research on the interplay between salience and informativeness* in attention to visual marketing stimuli, and on the role that consumer goals play in this process, is clearly needed.

Third, more work is welcome on other static visual marketing stimuli besides print ads, and on dynamic stimulus modes, such as television and Web sites and advertising. Whereas research to date has taught us much about visual attention to print ads, less is known about attention to other static visual marketing stimuli, such as product and brand packages, shelf design and layout, point-of-purchase material, outdoor communication. Also, next to nothing is known about attention to television commercials, product placement, in-script sponsoring, and similar dynamic visual marketing stimuli. Issues of the optimal length and pacing of television commercials, or optimizing brand exposures in them, are ready to be studied. Because of the "doubly dynamic" character of such data, this research will require major theoretical and data analytic efforts, but the rewards are likely to be high as well, in view of the dominant role of such dynamic stimuli in marketing practice. In a similar vein, more research on Web sites and advertising is awaited, which may even be more challenging, because static and dynamic modes are mixed in them, but the possible insights to be gleaned are likely to be equivalently rewarding as well. Such *widening the scope of the stimulus modes* under research scrutiny seems particularly fruitful.

Fourth, the time is ripe to develop and test standardized attention metrics to gauge the effects of visual marketing. A large set of eye-tracking measures has been used in previous marketing research, such as gaze duration, fixation duration, fixation frequency, first fixation, and the like. Originally such disaggregate, micro measures were often developed in the context of reading research, and they have been shown to be useful as well for visual marketing. Yet, to advance the application of eye-tracking research in marketing practice, it seems important to arrive at a smaller set of metrics—aggregated over consumers, time, and/or space—that tap key outcomes of the attention process of interest and that are attuned to the needs of visual marketing practice. Similar to marketing metrics such as conversion ratios or price elasticities, attention metrics might

become standards in the industry that marketers can compare to benchmarks in order to manage and assess their visual marketing efforts. We believe that the *development of standardized attention metrics* will accelerate the further development of visual marketing.

Finally, although there is evidence that even small attentional effects on commercial visual stimuli, such as print ads, catalogues, shelves, labels, and packaging designs, affect memory, attitudes, preferences, and intentions (Fletcher et al. 1995; Janiszewski 1998; Krugman et al. 1994; Pieters and Warlop 1999; Pieters, Warlop, and Wedel 2002; Rosbergen, Pieters, and Wedel 1997; Treistman and Gregg 1979; Wedel and Pieters 2000), more research is needed that investigates the relationship of attention to downstream marketing effects. Treistman and Gregg (1979) observed that one of two ads that people looked at longer in an eye-tracking experiment also received more sales. In view of the rapid increase of eye-tracking applications in practice and the increasing clutter in advertising practice, *investigating the role that visual attention plays in stimulating sales* would fill an important gap in current knowledge.

In conclusion, eye movements under normal conditions are strongly and directly connected to high-order cognitive processes, and attention plays a much more central role in communication processing and effectiveness than previously believed. Research to date has confirmed the effects of a host of bottom-up and top-down factors on the attention patterns to a range of commercial stimuli. The current opportunities to collect eye-tracking data on large samples of stimuli and consumers at comparatively low costs are unprecedented. Eye movements may contribute significantly to visual marketing reaching its full potential, by providing an exceptional view on consumers' moment-to-moment processing of visual marketing stimuli, with predictive validity for downstream effects. And there is now a substantial body of research for future visual marketing research to build on.

Acknowledgments

The authors thank Anocha Aribarg and Ulf Böckenholt for useful comments.

Notes

1. www.Tobii.com.
2. http://poynterextra.org/eyetrack2004/main.htm, last accessed April 3, 2006.

References

Anstis, S.M. 1974. "A Chart Demonstrating Variations in Acuity with Retinal Position." *Vision Research* 14: 589–592.

Aoki, Hirotaka, and Kenji Itoh. 2000. "Analysis of Cognitive Attitudes to Commercial Films on Basis of Eye-Tracking Data." *Proceedings of Human Factors and Ergonomic Society* 1: 38–41.

———. 2001. " Analysis of Cognitive Processes During Viewing Television Commercials Based on Semantic Structure of Scenes and Eye-Movement Data." *Journal of Japan Industrial Management Association* 52 (2): 101–116.

Biel, Alexander L. 1993. "Ad Research in the US: Hurdle or Help? A Brief Review of the State of the Art." *AdMap* 329 (May): 27–29.

Burke, R.R., and T.K. Srull. 1988. "Competitive Inference and Consumer Memory for Advertising." *Journal of Consumer Research* 15 (1): 55–68.

Buswell, G.T. 1935. *How People Look at Pictures: A Study of the Psychology of Perception in Art*. Chicago: University of Chicago Press.

Chandon, Pierre. 2002. "Do We Know What We Look At? An Eye-Tracking Study of Visual Attention and Memory for Brands at the Point of Purchase." Working Paper, INSEAD, Fontainebleau.

Cowen, Laura. 2001. "An Eye Movement Analysis of Web-Page Usability." M.A. thesis, Lancaster University, UK.

Davenport, Thomas H., and John C. Beck. 2001. *The Attention Economy: Understanding the Currency of Business*. Boston, MA: Harvard Business School Press.

Dreze, Xavier, and François-Xavier Hussherr. 2003. "Internet Advertising: Is Anybody Watching?" *Journal of Interactive Marketing* 17 (4): 8–23.

Duchowski, Andrew T. 2002. "A Breadth-First Survey of Eye-Tracking Applications." *Behavior Research Methods, Instruments, and Computers* 34 (4): 455–470.

———. 2003. *Eye-Tracking Methodology: Theory and Practice*. New York: Springer Verlag.

d'Ydewalle, G., G. Desmet, and J. Van Rensbergen. 1998. "Film Perception: The Processing of Film Cuts." In *Eye Guidance in Reading and Scene Perception*, ed. G. Underwood, 357–368. Amsterdam: Elsevier.

d'Ydewalle, G., and F. Tamsin. 1993. "On the Visual Processing and Memory of Incidental Information: Advertising Panels in Soccer Games." *Visual Search* 2: 401–408.

d'Ydewalle, G., P. Vanden Abeele, J. Van Ronsgerger, and D. Concke. 1988. "Incidental Processing of Advertisements While Watching Soccer Game Broadcasts." In *Practical Aspects of Memory: Current Research and Issues*, ed. M.M. Gruneberg, P.E. Morris, and R.N. Sykes, 470–483. Vol. 1. *Memory in Everyday Life*. Chichester, NY: Wiley.

Feng, Gary 2003. "From Eye Movement to Cognition: Toward a General Framework of Inference. Commentary on Liechty et al. 2003." *Psychometrika* 68: 551–556.

Fischer, Paul M. John W. Richards, Earl J. Berman, Dean M. Krugman. 1989. "Recall and Eye-Tracking Study of Adolescents Viewing Tobacco Advertisements." *Journal of the American Medical Association* 261 (1): 84–89.

Fitts, P.M., R.E. Jones, J.L. Milton. 1950. "Eye Movements of Aircraft Pilots during Instrument-Landing Approaches." *Aeronautical Engineering Review* 9 (2): 24–29.

Fletcher, James E., Dean M. Krugman, Richard J. Foit, Paul Fischer, and Tina Rojas. 1995. "Masked Recall and Eye-Tracking of Adolescents Exposed to Cautionary Notices in Magazine Ads." *Marketing and Public Policy Conference Proceedings*, ed. P.S. Ellen and P.J. Kaufman, 128–135. Vol. 5. Atlanta: Georgia State University Press.

Fox, Richard J., Dean M. Krugman, James E. Fletcher, and Paul M. Fischer. 1998. "Adolescents' Attention to Beer and Cigarette Print Ads and Associated Product Warnings." *Journal of Advertising* 27 (3): 57–68.

Goldberg, Joseph H., Claudia K. Probart, and R.E. Zak. 1999. "Visual Search of Food Nutrition Labels." *Human Factors* (September): 425–437.

Groves, Robert M., Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2004. *Survey Methodology*, Hoboken, NJ: Wiley-Interscience.

Harris, C.M. 1993. "On the Reversibility of Markov Scanning in Free Viewing." *Visual Search* 2: 123–135.

Henderson, John M., and Andrew Hollingworth 1998. "Eye Movements During Scene Viewing: An Overview." In *Eye Guidance in Reading and Scene Perception*, ed. Geoffrey Underwood, 269–293. Amsterdam: Elsevier.

Jacob, R.J.K., and K.S. Karn. 2003. "Eye-Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises." In *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, ed. R. Radach, J. Hyona, and H. Deubel, 573–605. Boston: North-Holland/Elsevier.

Janiszewski, Chris. 1998. "The Influence of Display Characteristics on Visual Exploratory Search Behavior." *Journal of Consumer Research* 25 (December): 290–301.

Josephson, Sheree, and Michael E. Holmes. 2002. "Visual Attention to Repeated Internet Images: Testing the Scanpath Theory on the World Wide Web." *Proceedings of ETRA 2002, Eye-Tracking Research & Applications Symposium*, 43–49.

Karslake, J.S. 1940. "The Purdue Eye-Camera: A Practical Apparatus for Studying the Attention-Value of Advertisements." *Journal of Applied Psychology* 24: 417–440.

Keller, Kevin L. 1991. "Memory and Evaluation Effects in Competitive Advertising Environments." *Journal of Consumer Research* 17 (March): 463–476.

Kent, Robert. 1993. "Competitive Versus Noncompetitive Clutter in Television Advertising." *Journal of Advertising Research* 33 (2): 40–46.

Kroeber-Riel, W. 1979. "Activation Research: Psychobiological Approaches in Consumer Research." *Journal of Consumer Research* 5 (March): 240–250.

Krugman, Dean M., R.J. Fox, J.E. Fletcher, P.M. Fischer, and T.H. Rojas. 1994. "Do Adolescents Attend to

Warnings in Cigarette Advertising? An Eye-Tracking Approach." *Journal of Advertising Research* 34: 39–52.
LaBerge, David. 1995. *Attentional Processing: The Brain's Art of Mindfulness*. Cambridge: Harvard University Press.
Laughery, Kenneth R. Stephen L. Young, Kent P. Vaubel, and John W. Brelsford, Jr. 1993. "The Noticeability of Warnings on Alcoholic Beverage Containers." *Journal of Public Policy and Marketing* 12 (1): 38–56.
Leven, Wilfried. 1991. *Blickverhalten von Konsumenten: Grundlagen, Messung und Anwendung in der Werbeforschung*. Heidelberg: Physica-Verlag.
Liechty, John, Rik Pieters, and Michel Wedel. 2003. "Global and Local Covert Visual Attention: Evidence from a Bayesian Hidden Markov Model." *Psychometrika* 68 (4): 519–541.
Lohse, Gerald L. 1997. "Consumer Eye Movement Patterns on Yellow Page Advertising." *Journal of Advertising* 26: 61–73.
Lohse, Gerald L., and D.J. Wu. 2001. "Eye Movement Patterns on Chinese Yellow Pages Advertising." *Electronic Markets* 11 (2): 87–96.
Mulvihill, Geoff. 2002. "Campbell Upgrades Soups to Battle Its Competitors." Available at www.freep.com/features/food/ffill31_20021231.htm.
Nixon, H.K. 1924. "Attention and Interest in Advertising." *Archives of Psychology* 72: 5–67.
Noton, D., and L. Stark. 1971. "Eye Movements and Visual Perception." *Scientific American* 224: 34–43.
Pieters, Rik, Edward Rosbergen, and Michel Wedel. 1999. "Visual Attention to Repeated Print Advertising: A Test of Scanpath Theory." *Journal of Marketing Research* 36 (November): 424–438.
Pieters, Rik, and Luk Warlop. 1999. "Visual Attention during Brand Choice: The Impact of Time Pressure and Task Motivation." *International Journal of Research in Marketing* 16: 1–17.
Pieters, Rik, Luk Warlop, and Michel Wedel. 2002. "Breaking Through the Clutter: Benefits of Advertisement Originality and Familiarity on Brand Attention and Memory." *Management Science* 48 (6): 765–781.
Pieters, Rik, and Michel Wedel. 2004. "Attention Capture and Transfer in Advertising: Brand, Pictorial and Text-Size Effects." *Journal of Marketing* 68 (April): 36–50.
Pieters, Rik, and Michel Wedel. 2007. "Goal Control of Attention to Advertising: The Yarbus Implication." *Journal of Consumer Research* 34 (August): 224–233.
Pieters, Rik, Michel Wedel, and Jie Zhang. 2007. "Optimal Feature Advertising Under Competitive Clutter." *Management Science*, forthcoming.
Poffenberger, Albert T. 1925. *Psychology in Advertising*. Chicago: A.W. Shaw.
Rayner, Keith. 1998. "Eye Movements in Reading and Information Processing: 20 Years of Research." *Psychological Bulletin* 124 (3): 372–422.
Rayner, Keith, Karen M. Rotello, Andrew J. Stewart, Jessica Keir, and Susan A. Duffy. 2001. "Integrating Text and Pictorial Information: Eye Movements When Looking at Print Advertisements." *Journal of Experimental Psychology: Applied* 7 (1): 219–226.
Redline, Cleo D., and Christopher P. Lankford. 2001. "Eye-Movement Analysis: A New Tool for Evaluating the Design of Visually Administered Instruments (Paper and Web)." Paper presented at the American Association for Public Opinion Research meeting, Montreal, Canada, May.
Reichle, E.D. and Nelson, J.R. 2003. "Local vs. global attention: Are two states necessary? Comment on Liechty et al. 2003." *Psychometrika*, *68*, 543–549.
Rimey, R.D., and C.M. Brown. 1991. "Controlling Eye Movements with Hidden Markov Models." *International Journal of Computer Vision* 7 (1): 47–65.
Rizzolatti, Giacomo, L. Riggio, and B.M. Sheliga. 1994. "Space and Selective Attention." In *Attention and Performance*, ed. C. Umiltà and M. Moscovitch, 231–265. Cambridge, MA: MIT Press.
Rosbergen, Edward, Rik Pieters, and Michel Wedel. 1997. "Visual Attention to Advertising: A Segment-Level Analysis." *Journal of Consumer Research* 24 (December): 305–314.
Russo, J. Edward. 1978. "Eye Fixations Can Save the World: A Critical Evaluation and a Comparison between Eye fixations and Other Information Processing Methodologies." In *Advances in Consumer Research*, ed. H.K. Hunt. Ann Arbor, MI: Association for Consumer Research, 561–570.
Russo, J. Edward, and France Leclerc. 1994. "An Eye-Fixation Analysis of Choice Processes for Consumer Nondurables." *Journal of Consumer Research* 21 (September): 274–290.
Russo, J. Edwards, and Larry D. Rosen. 1975. "An Eye Fixation Analysis of Multialternative Choice." *Memory and Cognition* 3 (3): 267–276.
Starch, Daniel. 1923. *Principles of Advertising*. Chicago: A.W. Shaw.

Stark, L.W., and H.C. Ellis. 1981. "Scanpaths Revisited: Cognitive Models Direct Active Looking." In *Eye Movements: Cognition and Visual Perception*, ed. D.F. Fisher, R.A. Monty, and J.W. Senders. Hillsdale: Lawrence Erlbaum Associates.

Stewart, A.J., M.J. Pickering, and P. Sturt. 2004. "Using Eye Movements During Reading as an Implicit Measure of the Acceptability of Brand Extensions." *Applied Cognitive Psychology* 18: 697–709.

Strong, Edward K. 1920. "Theories of Selling." *Journal of Applied Psychology* 9 (February): 75–86.

Treistman, Joan, and John P. Gregg. 1979. "Visual, Verbal, and Sales Responses to Print Ads." *Journal of Advertising Research* 19 (4): 41–47.

Ungerleider, L.G., and M. Mishkin. 1982. "Two Cortical Visual Systems." In *The Analysis of Visual Behavior*, ed. D. Ingle, R.J.W. Mansfeld, and M.S. Goodale, 549–586. Cambridge, MA: MIT Press.

van Raaij, Fred W. 1977. "Consumer Information Processing for Different Information Structures and Formats." In *Advances in Consumer Research*, ed. W.D. Perreault, Jr., 176–184. Vol. 4. Atlanta: Association for Consumer Research.

Wedel, Michel, and Rik Pieters. 2000. "Eye Fixations on Advertisements and Memory for Brands: A Model and Findings." *Marketing Science* 19 (4): 297–312.

Wedel, M., Pieters, R., & Liechty, J. (2003). "Evidence for covert attention switching from eye-movements; Reply to Commentaries on Liechty et al. 2003." *Psychometrika*, 68 (4), 557–562.

Witt, Dieter. 1977. "Emotional Advertising: The Relationship Between Eye-Movement Patterns and Memory—Empirical Study with the Eye-Movement Monitor." Ph.D., University of Saarland. Described in Kroeber-Riel 1979.

Yarbus, Alfred L. 1967. *Eye Movements and Vision*. New York: Plenum Press.

Young, L.R., and D. Sheena. 1975. "Survey of Eye Movement Recording Methods." *Behavioral Research Methods and Instrumentation* 7: 397–429.

CHAPTER 6

ROLE THEORY APPROACHES FOR EFFECTIVENESS OF MARKETING-ORIENTED BOUNDARY SPANNERS

Comparative Review, Configural Extension, and Potential Contributions

JAGDIP SINGH AND ARGUN SAATCIOGLU

Abstract

Role theory has proved remarkably promising in examining effectiveness of marketing-oriented boundary spanners. This paper reviews different approaches for examining role theory implications for boundary spanners—namely universalistic and contingency approaches—and develops the configural approach by extending configurational theory principles to role theory. Neither the contingency nor the configural approach has received much attention in the marketing literature. We compare and contrast different approaches, outlining bodies of work that have remained less accessible to marketing researchers. By triangulating across the alternative approaches, we expose underlying assumptions and press for critical assessment of their ecological validity. We identify opportunities for potential contributions by exploring promising but as yet uncharted approaches.

Marketing oriented boundary spanners such as salespeople, frontline, and customer contact employees fill critical roles that influence organizational effectiveness and sustainability. Consider the following:

- Boundary spanners are strategically important because they represent the "face" of the organization to customers and public, and are critical nodes where knowledge about markets and consumers is accumulated.
- Boundary spanning work is rarely routinized and involves significant people-oriented work. Boundary spanners are required to constantly interact with customers, undertake tasks that involve emotional labor, and provide discretion to tailor their behaviors to individual customer needs, problems, and demands.
- Boundary spanning work is sensitive to internal and external organizational environments. Variation in consumer demands (e.g., seasonal and/or economic variations in demand for products/services) and in internal operations (e.g., new product/service introductions or interface technologies) often affects boundary spanners unpredictably.

- Boundary spanners are organizationally monitored and controlled (e.g., via human supervision, electronic, audio, and video devices). Organizations are increasingly concerned about the productivity of boundary spanners, while keeping the quality of customer service delivered in focus.
- Boundary spanning work is highly stressful. Such work is likened to a "three-cornered fight" with the customer (demanding attention and service) and the organization (demanding efficiency and productivity) at the two ends and the boundary spanner "caught in the middle."
- Boundary spanning roles are profit centers. They are expected to cross-sell, up-sell, and more-sell while in the process of providing high-quality service/information. This dual accountability injects competing pressures on boundary spanners.

Of the various theories applied to study effectiveness of marketing-oriented boundary spanners, role theory is arguably the most promising so far. With its roots in sociology dating back to the fifties (Merton 1949; Rommetveit 1954), early grounded research on work organizations can be traced to the sixties. The much-cited work of Kahn et al. (1964) and Belasco's (1966) research with salespeople were important steps in translating sociological notions of role theory into meaningful and relevant constructs for the study of marketing-oriented boundary spanners. For instance, independent of Kahn et al. (1964), Belasco (1966) made some important observations on the different role demands experienced by salespeople: (a) intellectual demands that require intelligence, problem-solving skills, and job knowledge abilities, (b) emotional demands from dealing with issues such as "advocacy conflict"—internal conflict from being an advocate for the customer and the company at the same time, and (c) interactional demands that arise from the intensity and adaptability required in the diverse range of interactions. Without effective coping mechanisms, Belasco feared that salesperson effectiveness would be seriously undermined regardless of their intelligence, job knowledge, and/or skills. Although Kahn et al. and Belasco provided rich theory for probing boundary spanning roles, scientific progress lingered till Rizzo, House, and Lirtzman (1970) published validated scales of role conflict and ambiguity, and stirred up research interest in this topic. By 1985, Jackson and Schuler reported over 200 articles on role conflict and ambiguity in organizational settings that were published between 1970 and 1983; of these, 96 were original empirical studies that they meta-analyzed. A few years later, Brown and Peterson (1993) were able to locate 59 studies that focused specifically on *salesperson* role conflict and ambiguity and its influence on performance and satisfaction.

Despite this volume of research, academic and practitioner perspectives on boundary role stress are defined by convergence and contrasts. Both perspectives converge on the view that boundary role stressors incur heavy costs for the organization and individual alike because of lowered productivity, reduced motivation and commitment, and increased health costs (Cavanaugh et al. 2000; Maslach and Leiter 1997). In some professions, especially involving frontline and customer contact work, stressors have been described as reaching epidemic proportions (Marino 1997). Contrasting perspectives emerge when the influence of boundary role stressors is considered. In contrast to the convergence in the academic literature, practitioners have long argued about the potential of boundary role stressors to promote performance, enhance motivation, and spark creativity (Mohrman and Cohen 1995; Newton 1995). With regularity, the popular press has fancied workplace mantras such as, "it is better to burn out than to rust out," presumably to assure boundary spanners that they are not alone in facing stress and that stress can be turned into an opportunity to develop and enhance oneself. Paradoxically, while this notion of "eustress" has deep roots in the academic literature (Selye 1976; Yerkes and Dodson 1908), empirical studies have generally produced weak and mixed evidence. Thus, while much academic research suggests redesigning

and reconfiguring boundary work in a way that reduces, if not eliminates, critical role stressors (Tubre and Collins 2000), practitioners view such recommendations with little relevance since they hold that the nature of customer interface (e.g., people-oriented, nonroutinized work) and its unpredictability (e.g., variability in internal and external conditions) make role stress an inherent aspect of boundary roles.

To bridge these perspectives, a promising approach has been proposed by Karasek (1979) and his colleagues that views a singular focus on role stressors as myopic and misguided. Instead, it argues that the study of role stressors must simultaneously consider the job scope—the degree of autonomy, feedback, and participation afforded to boundary role employees (Karasek 1979; Xie and Johns 1995). Noting that greater job scope may make all the difference between "eustress" and "distress," and between "healthy" and "unhealthy" work, this approach is theoretically appealing because of its conceptual richness, and managerially attractive as evidenced by the popularity of empowerment programs. Unfortunately, such stressor–scope models have received limited empirical attention in the marketing literature. As such, the potential of the stressor–scope framework to provide insights and bridge perspectives is largely unrealized.

The purpose of this review is to illuminate and strengthen the preceding bridge to germinate new research ideas and directions for understanding the effectiveness of marketing-oriented boundary spanners. Specifically, we provide a review of three different theoretical perspectives on boundary role stress and effectiveness, including universalistic, contingency, and configural perspectives. The universalistic perspective reflects much current research in marketing and is grounded in the role episode model of Kahn et al. (1964). The contingency perspective is based on contributions to Karasek's model. Because many of these contributions have occurred outside the marketing literature, we provide a detailed discussion of the theory underlying this perspective, and review the associated empirical literature. Finally, we develop the configural perspective as a theoretical contribution of this paper. This perspective extends ideas from configurational approaches to posit nonlinear and higher-order effects of role theory that cannot be represented by contingency approaches. While these perspectives have competing elements, our orientation is comparative and complementary. We compare these perspectives for their theoretical distinctiveness to encourage future research that approaches the study of boundary role stress and effectiveness from multiple, not singular, perspectives to uncover convergent and anomalous ideas. Focusing on a singular theoretical perspective, as reflected in much marketing literature, limits the vision of understanding. In addition, at their boundaries, these theoretical perspectives offer opportunities for interesting and creative work that has remained as yet untapped and unexploited. We provide an outline for future research directions to this end.

Theoretical Perspectives on Role Stressors and Boundary Spanning Roles

Universalistic Perspective

This perspective posits that role stressors invariably have dysfunctional consequences for boundary spanner outcomes including performance, satisfaction, and commitment, regardless of job context, scope, and/or nature of the organization.

Theory

Rooted in Kahn et al.'s work, this perspective posits that role stressors have a linear relationship with boundary spanner outcomes. The commonly examined role stressors include role conflict,

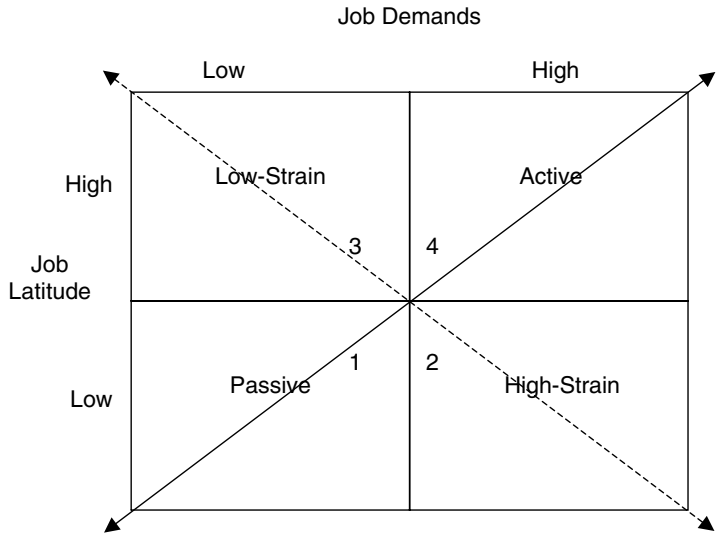
role ambiguity, and role overload (Behrman and Perreault 1984; Belasco 1966; Kahn et al. 1964). *Role conflict* occurs when a boundary spanner believes that the expectations and demands of two or more members of his or her role set are incompatible (e.g., boss and customer). *Role ambiguity* relates to the perceived lack of information needed by an employee to perform his or her role adequately and his or her uncertainty about the expectations of different role set members. *Role overload* occurs when the frontline employee perceives that the cumulative role demands exceed his or her abilities and motivation to perform a task. The influence of different role stressors on boundary spanner performance and well-being is supported conceptually by the role episode model of Kahn and colleagues (1964), which posits that (1) boundary spanners interact with different role senders (e.g., customers, boss, co-workers) in many episodes to obtain information, direction, task demands, and assistance; (2) role sender demands and expectations take the form of perceived stressors when a boundary spanner believes that there is conflict (e.g., among demands), ambiguity (e.g., about expectations), or overload (e.g., of demands and expectations); (3) perceived stressors are influenced by a person's psychological, dispositional, and sociological characteristics; and (4) persistent stressors are likely to overwhelm the person's resources and thereby have a dysfunctional impact on his or her behavioral and psychological outcomes (e.g., job performance, satisfaction). Hobfoll and Freedy (1993) have conceptualized the influence of role stressors within a conservation of resources framework. Boundary spanners are thought to regulate their behaviors to cope with role stressors in a way that conserves their valued resources; however, regulation failures occur when stressors overwhelm an individual's coping resources, resulting in impaired performance and well-being. While it allows for the possibility that different boundary spanners may perceive disparate levels of role stressors in similar work situations, this perspective is universalistic in its predictions about the linear and direct effect of perceived role stressors on outcomes.

Empirical Findings and Assessment

This perspective has produced significant empirical work summarized in several meta-analyses and reviews (Brown and Peterson 1993; Fisher and Gitelson 1983; Jackson and Schuler 1985; King and King 1990). The preponderance of evidence suggests that the influence of role stressors is consistent, compelling, and invariably dysfunctional (Behrman and Perreault 1984; Brown and Peterson 1993; Fisher and Gitelson 1983; Rhoads, Singh, and Goodell 1994; Singh 1993; Spector, Dwyer, and Jex 1988). For example, Brown and Peterson (1993), in their meta-study, found correlations of –.24, –.36, –.28, and .36 between role ambiguity and job performance, satisfaction, commitment, and propensity to leave, respectively. Likewise, the correlations between role conflict and job performance, satisfaction, commitment, and propensity to leave were –.07, –.33, –.34, and .28 respectively. Although role overload is not as frequently studied as other role stressors, in general the correlations between role overload and the different job outcomes parallel those obtained for role conflict and ambiguity in terms of both magnitude and direction. For instance, Singh, Goolsby, and Rhoads (1994) report correlations of –.14, –.39, .25, and .09 between role overload and job satisfaction, organizational commitment, turnover intentions, and job performance, respectively. As such, there is broad evidence supported by meta-analytical results that role stressors have significant linear and dysfunctional relationships with critical job outcomes (Brown and Peterson 1993; Fisher and Gitelson 1983; Jackson and Schuler 1985).

While researchers have called for exploring moderating variables, the majority of the empirical work has downplayed the effect of situational or contextual variability. For instance, in Brown and Peterson's meta-analysis, the influence of supervisory behaviors and job/task variables accounted for less than 10 percent of explained variance. Consequently, Brown and Peterson 1993,

Figure 6.1 **A Graphical Representation of Karasek's Job Demands (Role Stressors)–Decision Latitude Model**



(p. 68) claimed that their role stress model that *excludes* job context variables indicates "considerable robustness and generalizability . . . across relationships and study contexts." Likewise, in another meta-analysis, Churchill et al. (1985, p. 109) found that, on average, organizational context variables explain "only 1% of the variation in performance" and that this influence was the "lowest . . . among the six categories of predictors studied." These findings parallel Jackson and Schuler's (1985) meta-analysis in that contextual variables have significant, negative but rather weak (< 15 percent shared variance) direct effects on role stressors, with marginal direct effects on job outcomes.

Contingency Perspective

In accord with this perspective, the effect of role stressors on boundary spanner outcomes is contingent on a third variable, such that this effect may be dysfunctional, neutral, or even functional, depending on the level of the contingent factor. Within this perspective, different theoretical models have been proposed that specify the contingent variable and the mechanism involved in moderating the effect of role stressors. The most commonly used theoretical frameworks are rooted in Karasek's (1979) demand–latitude model displayed in Figure 6.1.

Theory

In Karasek's model, the influence of psychological demands—or role stressors in the context of boundary spanners—on job outcomes is contingent on the availability of decision latitude to the individual (e.g., autonomy)—also referred to as job scope or job control. Karasek suggests that certain modalities of the demand–latitude interplay result in higher job outcomes than do other modalities. In particular, Karasek asserts that increasing levels of psychological job demands must be matched with increasing levels of decision latitude for maintaining or enhancing performance and psychological well-being. Utilizing "low" and "high" distinctions for demands and latitude (see

Figure 6.1), Karasek developed the logic for the underlying mechanism for the differing influence of role stressors. When latitude is low and boundary spanners have little discretion in making task decisions, Karasek hypothesized that the job context would be either "passive" or "high-strain" corresponding to a "low" or "high" level of role stressors, respectively. Specifically, when role stressors are high, a "high-strain" work context is obtained because employees lack the necessary coping resources that come from decision latitude to deal with high levels of role demands, and are easily overwhelmed (see cell 2 in Figure 6.1). In a high-strain environment, boundary spanner performance is impaired and well-being undermined. By contrast, when role stressors are at a low level, with decision latitude also at a low level, Karasek suggested that a passive work context is obtained where both performance and well-being are suboptimal (see cell 1 in Figure 6.1). *Why so?* Drawing from alienation theory (Blood and Hulin 1967), Karasek reasoned that employees generally lack the stimulation to actively engage in tasks when role stressors are low. This passive orientation toward tasks is exacerbated by low levels of autonomy that fail to provide a sense of control over decisions that affect an individual's job, thereby inhibiting employee efforts to insert challenge and growth in low-stress jobs. Consequently, boundary spanning jobs with low levels of stressors and latitude are posited to be passive and suboptimal.

Now consider the contingencies when decision latitude is high (see cells 3 and 4 in Figure 6.1). Karasek predicted that the job context would be either "active" or "low-strain" corresponding to a high or low level of role stressors respectively. Specifically, when role stressors are low, Karasek reasoned that boundary spanners possess significant resources that stem from autonomy that can be deployed to address challenging job demands. However, the low level of stressors offers little by way of challenges to channel individual resources. This abundance of underutilized resources makes for a low-strain job context. Karasek did not view such low-strain jobs favorably. Rather, he argued that such job contexts lack the potential to grow individual skills and enhance self-efficacy and, consequently, are suboptimal (cell 3, Figure 6.1). By contrast, Karasek posited a favorable perspective for job contexts with high levels of role stressors. Building from motivation theories (Csikszentmihalyi 1975; Hackman and Oldham 1976), Karasek noted that employees with high levels of task control are likely to enjoy resources needed to cope with challenging demands. When the job context supplies these demands under conditions of "high" role stressors, the boundary spanner can draw from available resources to effectively cope with the challenge. Because effectively dealing with challenging job demands is efficacious, the boundary spanner is likely to grow from this experience by being more self-confident, resourceful, and energized to tackle future challenges (Csikszentmihalyi 1975). Karasek noted that such job contexts are "active" (cell 4, Figure 6.1).

Empirical Findings

A distinctive aspect of Karasek's theory is that it depicts "interaction effects" of role stressors and decision latitude on job outcomes. As such, in most cases, Karasek's hypothesis is tested by examining the significance of the interaction term involving appropriate measures for job demands and latitude, with performance or other job outcomes as the dependent variable. In empirical tests of Karasek's (1979) theory across a wide range of work contexts using a variety of designs (see Table 6.1 for a summary), the results for the interaction effects have been mixed. For instance, studies by Abdel-Halim (1981), Daniels and Guppy (1994), Kelloway and Barling (1991), and Landsbergis (1988), both validated and extended Karasek's model. Landsbergis (1988) found that job strain (e.g., dissatisfaction, depression, and psychosomatic symptoms) and burnout were higher in health care jobs that combined high job (workload) demands and job control (decision latitude). Likewise,

(text continues on page 172)

Table 6.1

A Summary of Key Empirical Studies Using a Contingency Perspective for Examining Role Theory Effects for Boundary Spanners

Authors	Model/Variables Utilized	Methodological Approach	Sample	Findings
Abdel-Halim (1978)	Dependent variables: Job satisfaction Job involvement Job anxiety Independent variables: Skill variety, task identity, task significance, autonomy, and feedback from job as moderating job enrichment characteristics combined into Motivating Potential Score (MPS) Role conflict, role ambiguity, and role overload as three distinct measures of role stress	Multiple linear regression analysis using interaction terms	89 randomly selected line and staff managers from six units of a large heavy-equipment manufacturing company	Interactions of role ambiguity and MPS ($F = 5.33$, $p <$. 05) and of role overload and MPS ($F = 10.93$, $p <$ .01) significantly predicted satisfaction ($R^2 = .44$, $F = 5.57$, $p < .05$). In low MPS situations (i.e., low job enrichment), increasing levels of role ambiguity and role overload reduced satisfaction. In high MPS situations, role ambiguity slightly reduced satisfaction whereas role overload considerably increased satisfaction. The interaction of role ambiguity with MPS ($F = 5.25$, $p < .05$) significantly predicted job involvement ($R^2 = .24$, $F =$ 5.20, $p < .05$). Role ambiguity decreased involvement in both low and high MPS situations, but much less in the latter than in the former. Finally, the interaction of role overload and MPS ($F = 6.62$, $p < .05$) significantly predicted job anxiety ($R^2 = .26$, $F = 6.57$, $p < .05$). Role overload increased anxiety in low MPS situations but reduced it in high MPS situations.

Abdel-Halim (1981)	Dependent variable: Satisfaction with work Independent variables: Role conflict, role ambiguity, and role overload as three distinct measures of role stress Skill variety, task identity, task significance, autonomy, and feedback as five distinct measures of job design characteristics/job complexity Organizational technology (nominally measured in two categories—long-linked manufacturing technology or mediating service technology)	Three-way analysis of variance analysis with main, two-way, and three-way interaction effects	170 randomly selected managerial and non-managerial personnel—89 from a heavy equipment manufacturing firm and 81 from five small to medium size banks	Three-way interactions of role ambiguity, job characteristics, and technology ($F = 5.43$, $p \leq .02$) and role overload, job characteristics, and technology ($F = 5.73$, $p \leq .01$) were statistically significant. In simple, low-scope jobs in long-linked manufacturing technologies and in complex, high-scope jobs in mediating service technologies, the interaction of two role stressors (ambiguity and overload) and job characteristics were found to be significant. The three-way interactions involving role conflict, and all two-way interactions were nonsignificant. These results indicate that organizational technology determines the manner in which job characteristics and role stressors interact.
Payne and Fletcher (1983)	Karasek's (1979) demand-control (discretion) model Dependent variables: Satisfaction with work, headmaster/headmistress, colleagues, and with the felt pressure on the job Felt adequacy of pay Depression Anxiety Obsession Somatic complaints	Cluster analysis using Ward's method, ANOVA, and multiple regression	148 teachers from various towns in the Midlands of England	Cluster analysis based on the independent variables identified seven groups with distinct combinations of demands, control, and support. ANOVA across the seven groups indicated that there were significant differences on satisfaction with the headmaster, colleagues, pay, and the felt pressure on the job. Group 2 (low disciplinary demands, average on other demands, high support, and high discretion) was significantly ($p < .05$) more satisfied than all other groups. Overall, the demands, supports, discretion model accounted for a small proportion of the variance in the dependent variables (between 10% and 21% of the satisfaction measures). The three-way

(*continued*)

Table 6.1 *(continued)*

Authors	Model/Variables Utilized	Methodological Approach	Sample	Findings
	Cognitive impairments Independent variables: Disciplinary demands, maintaining standards, and workload demands as three distinct measures of job demands Interpersonal support Job discretion			interaction of demands, discretion, and support were not included in the multiple regression analysis. The demand/discretion interaction (as originally proposed by Karasek) did not significantly predict any of the dependent variables.
Spector (1987)	Karasek's (1979) demand-control (discretion) model Dependent variables: Total satisfaction and general satisfaction as two distinct measures of satisfaction Frustration Anxiety Health symptoms (having them and going to the doctor for these symptoms as two distinct measures) Independent variables: Discretion (control) Workload, role ambiguity, role conflict, and interpersonal conflict as distinct measures of job demands; sum of the standardized scores of these measures as a combined measure of job demands	Multiple regression analysis, using interaction terms	136 clerical employees drawn from different positions at a major state university	Each dependent variable was regressed on control, a stressor, and the interaction of control and that stressor. Out of a total of 30 regression analysis, only 2 (7%) had interaction terms that were significant, providing weak support to Karasek's original interaction hypothesis. The form of these interactions were found not to correspond to those required by Karasek's hypothesis either.

Landsbergis (1988)	Karasek's (1979) demand-control model <u>Dependent variables:</u> Job dissatisfaction, sleeping problems, depression/life dissatisfaction, and physical/psychosomatic strain as four distinct measures of psychological strain Coronary heart disease (CHD) symptom indicator Frequency and intensity of emotional exhaustion, depersonalization, and personal accomplishment as six distinct measures of burnout measured by Maslach Burnout Inventory (MBI) <u>Independent variables:</u> *Job characteristics (JCS Survey)* Skill discretion, decision authority, union influence, and participation as four distinct measures of decision latitude Workload demands Supervisor support and co-worker support as two distinct measures of social support Job insecurity	MANOVA, ANCOVA (Roy/Bargmann step-down analysis), hierarchical multiple regression using interaction terms, canonical correlations, and t-tests Step 1: Examining mean differences of dependent variables across Karasek's original four configurations of demands and latitude using MANOVA and ANCOVA procedures Step 2: Predicting strain and other outcomes with workload demands, decision latitude, demand–latitude interaction, social support, job insecurity, physical exertion, and hazard exposure Step 3: Canonical correlation with demographics and job characteristics as independent variables, and the strain scales, CHD indicator, smoking, and burnout measures as dependent variables	771 employees composed of 15 different job titles in 4 bargaining units in 2 hospitals and 1 nursing home	At Step 1, MANOVA indicated that, across all the dependent measures, there were significant differences ($p <$ .001, df = 33/690) between the four quadrants of Karasek's original model. ANCOVA indicated that job dissatisfaction, emotional exhaustion, and personal accomplishment differed across the four quadrants ($p < .05$; F-values ranged from 3.11 to 28.87; df ranged from 3/234 to 3/244). In particular, job dissatisfaction and emotional exhaustion were significantly higher among high-strain employees, and the active and passive quadrants did not differ in terms of psychological strain measures—but active employees did report higher emotional exhaustion than did passive ones. The active quadrant was also characterized by high personal accomplishment compared to all other quadrants. At Step 2, hierarchical regression indicated that the demand–latitude interaction did not significantly predict any of the four measures of psychological strain. The main effects of demand and latitude, as well as of social support, hazard exposure, physical exertion, and job security significantly ($p < .05$) predicted all of the four strain measures and accounted for significant portions of variation. For job satisfaction, ΔR^2 ranged from .046 to .152 and the resulting R^2 was .433 (F = 6.28, $p < .001$). For depression, ΔR^2

(*continued*)

Table 6.1 *(continued)*

Authors	Model/Variables Utilized	Methodological Approach	Sample	Findings
	Physical exertion Hazard exposure *Job Strain (JCS Survey)* Depression/life dissatisfaction Job dissatisfaction Physical/psychosomatic strain Sleeping problems			ranged from .020 to .065 and the resulting R^2 was .163 ($F = 4.22$, $p < .01$). For physical strain, ΔR^2 ranged from .018 to .073, the resulting R^2 was .195 ($F = 4.88$, $p < .001$). For sleeping problems, ΔR^2 ranged from .011 to .046, the resulting R^2 was .133 ($F = 2.58$, $p < .05$). The R^2 for CHD was .13 and the demand–latitude interaction did significantly increase the R^2. Similar results were obtained for the burnout scales as dependent variables. The demand–latitude interaction did not significantly predict any of the strain measures. Workload demands significantly increased R^2 for emotional exhaustion and depersonalization but not personal accomplishment. Latitude increased R^2 for all burnout scales, and union activity significantly increased R^2 except for personal accomplishment. The resulting R^2s for emotional exhaustion, depersonalization, and personal accomplishment were .40, .22 to .23, and .25 to .38, respectively.
Perrewe and Ganster (1989)	Karasek's (1979) demand-control model, and a workload-control model <u>Dependent variables:</u> Task satisfaction, measured using GM Faces Scale and Semantic Differential scale Perceived anxiety measured using Affect Adjective Check List (AACL) and	Multiple regression analysis using interaction terms	125 volunteer undergraduate students enrolled in an introductory management course	Objective demands had a significant main effect only on satisfaction ($\beta = -.94$, $F = 14.52$, $p < .001$, $R^2 = .12$) using only the Faces scale. Objective control had no significant main effects on any of the dependent variables. Perceived demands had significant negative main effects on satisfaction using both the Semantic Differential scale ($\beta = -.29$, $F = 9.46$, $p < .01$, $R^2 =$

Authors	Model/Variables Utilized	Methodological Approach	Sample	Findings
	Subjective Stress scale Pulse rate and skin temperature as two distinct measures of psychological arousal <u>Independent variables:</u> Objective and perceived behavioral control Objective and perceived workload demands			.08) and the Faces scale (β = –.50, F = 24.45, $p < .001$, $R^2 = .18$). Perceived demands also had significant positive effects on perceived anxiety using both the Subjective Stress scale (β = .20, F = 15.56, $p < .001$, $R^2 = .13$) and AACL (β = .52, $F = 45.26$, $p < .001$, $R^2 = .29$). Perceived control had a significant main effect only on anxiety using the AACL (β = –.02, $F = 4.06$, $p < .001$, $\Delta R^2 = .03$, $R^2 = .32$). The findings on the main effects of demand and control were largely consistent with Karasek's original postulations. As far as the interaction effects were concerned, only the perceived demand/perceived control interaction was found to have a significant effect. This effect was on perceived anxiety (β = –.07, $F = 2.87$, $p < .10$, $\Delta R^2 = .02$, $R^2 = .17$). Controlling for the level of perceived control (low, medium, high), it was found that the relationship between perceived demands and perceived anxiety became weaker as perceived control got higher.

(*continued*)

Table 6.1 *(continued)*

Authors	Model/Variables Utilized	Methodological Approach	Sample	Findings
Warr (1990)	Karasek's (1979) demand-control (decision latitude) model and Warr's (1987) two-dimensional affective well-being model that includes anxiety, depression, pleasure, and arousal Dependent variables: Anxiety-contentment, depression-enthusiasm, pleasure-displeasure (job satisfaction) as three distinct measure of affective well-being Independent variables: Job demands Decision latitude	Multiple regression analyses using quadratic and interaction terms	1,686 employees of research company in 75 locations in the UK	Both job demands and decision latitude were found to have significant linear main effects on all three measures of well-being (βs ranged from .60 to 1.37, $p < .01$), except for the effect of decision latitude on anxiety-contentment. Job demands was found to have a nonlinear (quadratic) effect on job satisfaction (β = –1.28, significance of $\Delta R^2 < .01$, $R^2 = .19$), anxiety-contentment (β = –.99, significance of $\Delta R^2 < .01$, $R^2 = .29$), and depression-enthusiasm (β = –.85, significance of $\Delta R^2 < .01$, $R^2 = .12$). Decision latitude had a nonlinear effect only on job satisfaction (β = –.46, significance of $\Delta R^2 < .01$, $R^2 = .42$). Finally, the interaction of job demands and decision latitude did not have a significant effect on any of the three measures of well-being. These results provide substantial support for Karasek's original ideas and findings regarding the nonlinear effects of demand and latitude on key job outcomes. However, they provide no support for his hypothesis regarding the interactive effects of demands and control.

Dwyer and Ganster (1991)	Karasek's (1979) demand-control model Dependent variables: Absence, tardiness, and sick days as three distinct measures of employee withdrawal Job satisfaction (overall) and work satisfaction (aspects of the work) as two distinct measures of satisfaction Independent variables: Workload demands (subjective), mental demands (objective), and physical demands (objective) as three distinct job demands Control (subjective)	Multiple regression analysis, using interaction terms	115 full-time workers (blue-collar and trade occupations) in a large manufacturing plant	Four significant interaction effects were found. The interaction between control and perceived workload demands had a negative effect on absence ($\beta = -1.00$, $F_{(1,86)} = 16.12$, $p < .01$, $\Delta R^2 = .15$) and a positive effect on satisfaction ($\beta = .43$, $F_{(1,86)} = 4.05$, $p < .05$, $\Delta R^2 = .04$), indicating that at high levels of control, perceived workload is associated with low absenteeism and high work satisfaction. Under low conditions of control, however, workload tends to increase absenteeism and decrease satisfaction. The interaction between control and objective mental demands had a negative effect on both tardiness ($\beta = -3.50$, $F_{(1,86)} = 31.49$, $p < .01$, $\Delta R^2 = .26$) and sick days ($\beta = -3.98$, $F_{(1,86)} = 4.62$, $p < .05$, $\Delta R^2 = .04$), indicating that at high levels of control, mental demands are associated with lower levels of tardiness and sick days. When control is low, mental demands lead to high tardiness and sick days. The interaction of control and physical demands did not have a significant effect on any of the dependent variables.

(*continued*)

Table 6.1 *(continued)*

Authors	Model/Variables Utilized	Methodological Approach	Sample	Findings
Kelloway and Barling (1991)	Warr's (1987) model of employment and mental health, which distinguishes between job-related and context-free mental health Dependent variable: Context-free mental health (General Health Questionnaire-12 unidimensional measure) Job-related subjective competence (personal accomplishments at work) as mediating variable Work satisfaction, emotional exhaustion, depersonalization as three distinct mediating measures of job-related affective well-being Independent variables: Autonomy, task variety, task identity, feedback from job, feedback from co-workers as five distinct measures of job characteristics Role ambiguity and role conflict as two distinct measures of role stress	Structural equation modeling. A processual model where subjective competence and affective well-being mediate the effects of job characteristics and role stressors on context-free mental health	All 2,300 employees of a large hospital were included in the study. 720 usable responses were obtained. The nursing staff made up 43% of the group	The structural equation model fit the data well ($Q = 84$, $W_{19,720} = 35.01$, $p < .01$). Context-free mental health was significantly predicted ($p < .01$) by subjective competence (–.14). It was also significantly predicted ($p < .001$) by all three dimensions of affective well-being—satisfaction (–.22), exhaustion (.37), depersonalization (.09). Subjective competence was significantly predicted ($p < .001$) by three job characteristics—variety (.27), autonomy (.23), feedback from job (.13). Dimensions of job-related well-being were significantly predicted ($p < .001$) by both role stressors: role ambiguity→ satisfaction, exhaustion, depersonalization (–.17, .10, .25); role conflict → satisfaction, exhaustion, depersonalization (–.32, .13, –.24). In addition, satisfaction was significantly predicted ($p < .01$) by autonomy (.17), task variety (.37), and feedback from co-workers (.15). These results not only confirm Warr's (1987) original contention that job and role characteristics predict job-related mental health, but also extend Warr's propositions by demonstrating that job-related mental health mediates the effects of role stressors and job characteristics on context-free mental health.

Fletcher and Jones (1993)	Karasek's (1979) demands-discretion (control) model. <u>Dependent variables:</u> Job satisfaction Life satisfaction Anxiety Depression <u>Independent variables:</u> Job demands (combination of pace of work, workload, job overload, and conflicting demands) Job discretion (combination of skill discretion and decision authority) Interpersonal supports (combination of professional competence of colleagues and support from boss, co-workers, and spouse/partner/family)	Multiple regression analysis using interaction and quadratic terms (for effects analysis), and MANOVA (for means analysis) Step 1: Replicating Karasek's model using multiple regression with main effects (three dependent variables: depression, job satisfaction, life satisfaction) Step 2: Testing Karasek's model for occupational status using MANOVA Step 3: Testing Karasek's model using multiple regression with interaction terms and controlling for occupation status Step 4: Testing curvilinear relationships using hierarchical multiple regression analysis Step 5: Testing for the effect of interpersonal support using hierarchical multiple regression	3,086 patients between the ages of 30 and 60 (1678 women, 1408 men) on the list of an NHS Health Center Practice in Hertfordshire, UK	MANOVA indicated that, for both men and women, there were significant differences ($p < .01$) in total demand and discretion scores of the two occupational groups (manual and nonmanual workers), both being significantly higher for the nonmanual workers (total differences ranged from 1.2 to 4.5; F-values ranged from 46.2 to 285.6). Multiple regressions controlling for occupational status indicated that main effects of demands and discretion were significant ($p < .05$) but again with weak explanatory power (R^2s ranged from .027 to .045). A significant effect ($p < .05$) was obtained only on job satisfaction ($\beta = .46$) and life satisfaction for men, but no significant effect on any of the dependent variables in the case of women. In general, for both men and women, the amount of variance in each dependent variable accounted for by the independent variables was very small (R^2s ranged from .018 to .095). There was no evidence of interaction and little evidence of curvilinear relationships.

(*continued*)

Table 6.1 *(continued)*

Authors	Model/Variables Utilized	Methodological Approach	Sample	Findings
Fox, Dwyer, Ganster (1993)	Karasek's (1979) demands-control model Depenedent variables: Job performance (combination of patient assessment, planning, developing patient care plans, etc.) Job satisfaction and complaints (illness and somatic) as two distinct measures of affective outcomes Arterial systolic and diastolic blood pressure at home and at work, and salivary cortisol at home and at work) as eight distinct measures of physiological outcomes (physiological data was collected after questionnaires were completed) Independent variables: Patient load, patient contact hours as percentage of total work, and number of deaths witnessed as three distinct objective measures of work demands Work load and inventory of stressful events as two distinct subjective measures of work demands Control (control over task variety, task order, pacing, scheduling, procedures, etc.)	Hierarchical multiple regression analysis with interaction terms	All of the 198 nurses employed in a medium-sized private hospital	The interaction of perceived control and perceived workload significantly predicted job satisfaction, systolic blood pressure at home and at work, diastolic blood pressure at home, and cortisol level at work (R^2s ranged from .11 to .59, F-values ranged from 4.43, df = 1128, $p < .05$ to 6.29, df = 1128, $p <$ .05). In low-control situations, workload reduced satisfaction, increased systolic blood pressure at work and at home, and cortisol levels at work. In high-control situations, workload increased satisfaction, reduced systolic blood pressure at work, and diastolic blood pressure at home, and did not affect systolic blood pressure at home and cortisol levels at work. The interactions of perceived control and objective work load significantly predicted job satisfaction and cortisol levels at home (R^2s ranged from .13 to .24, F-values ranged from 3.98, df = 1146, $p < .05$ to 8.05, df = 1147, $p <$ .01).Two of these interactions, control/stressful events and control/patient contact time, support the Karasek model. In low-control situations, stress events reduced satisfaction and patient contact time increased cortisol levels at home. In high-control situations, stress events did not affect satisfaction and patient contact time only slightly reduced cortisol levels at home.

Daniels and Guppy (1994)	Warr's (1987) model of employment and mental health, which distinguishes between job-related and context-free mental health <u>Dependent variables:</u> Job-related pleasure-displeasure, job-related anxiety-contentment, and job-related depression-enthusiasm as three distinct context-dependent measures of psychological well-being General Health Questionnaire-12 as the context-free unidimensional measure of well-being <u>Independent variables:</u> Work locus of control (WCLS) Job stressors Social support Job autonomy (combination of task independence, closeness of supervision) Participation in decision making	Longitudinal repeated measures design using multiple moderated regression analysis with interaction terms All constructs measured twice with a month gap Model 1: with participation Model 2: with autonomy Model 3: with social support In all models: Step 1: initial well-being scores Step 2: main effects Step 3: all two-way interactions Step 4: three-way interactions	244 randomly selected accountants from the Institute of Chartered Accountants of England and Wales	No significant two-way interactions were obtained for participation/stressors, autonomy/stressors, and social support/stressors. Significant three-way interactions obtained for stressors/WCLS/autonomy ($p < .05$) and for stressors/WCLS/social support ($p < .10$). The significant three-way interactions indicate that WCLS and job autonomy buffer the effects of stressors on psychological well-being such that stress compromises well-being the least when locus of control is internal and autonomy is high, and most when locus of control is external and autonomy is low.

(*continued*)

Table 6.1 *(continued)*

Authors	Model/Variables Utilized	Methodological Approach	Sample	Findings
Xie and Johns (1995)	Dependent variables: Emotional exhaustion (chronic stress) and anxiety (long-term stress) as two distinct measures of stress Independent variables: Job complexity from Dictionary of Occupational Titles (DOT) and Occupational Prestige Index (OP) as two distinct objective measures of job scope Subjective job scope (combination of skill variety, task identity, task significance, autonomy, and feedback) Perceived demands-ability fit	Curvilinear hierarchical moderated multiple regression analysis using interaction terms	415 full-time employees (professionals, managers, sales workers, clerical workers, blue-collar workers) from a random selection of 65 organizations including banks, insurance companies, high schools, hospitals, etc.	Subjective job scope, DOT, OP, task significance, and task identity significantly predicted exhaustion in a U-shaped curvilinear fashion (ΔR^2s ranged from .01 [R^2 = .11] to .04 [R^2 = .11], F-values ranged from $F_{1,403}$ = 4.55, $p < .05$ to $F_{1,409}$ = 19.09, $p < .001$). Low and high levels of these predictors were associated with high levels of exhaustion, whereas at moderate levels they predicted low levels of exhaustion. Anxiety was significantly predicted in U-shaped curvilinear fashion only by task identity (ΔR^2 = .01, [R^2 = .03], $F_{1,400}$ = 5.08, $p < .05$), indicating that low and high levels of identity were associated with high anxiety whereas moderate levels of identity predicted low anxiety. DOT/demand-ability fit and OP/demand-ability fit interactions significantly predicted both exhaustion and anxiety (ΔR^2s ranged from .009 [R^2 = .09] to .015 [R^2 = .10], F-values ranged from $F_{1,409}$ = 4.03, $p < .05$ to $F_{1,401}$ = 6.86, $p <$ .01). These interactions were slightly and negatively associated with exhaustion and anxiety.
Wall, Jackson, Mullarkey, and Parker (1996)	Karasek's (1979) demand-control model, and a demand–latitude model Dependent variables:	Moderated regression analysis using interaction terms	All 1,451 production employees and support and supervisory staff	The interaction effects of job demands and job control were significant in predicting each dependent variable. F-values ranged from 7.81 to 13.04

	Job satisfaction, depression, and anxiety as three distinct measures of job strain Independent variables: Job demands (combination of monitoring demand and problem-solving demand) Job control (combination of timing control and method control) Decision latitude (Perceived Intrinsic Job Characteristics Scale, which includes measures of freedom to choose method, opportunity to use ability, amount of variety, etc.)		from four British manufacturing companies	($p < .01$). The interaction effects of job demands and decision latitude were, however, nonsignificant in predicting the dependent variables. F-values ranged from .95 to 2.37 ($p < .01$). Results indicate that (psychological) job strain is caused by the interaction effect of job demands and job control, not of job demands and decision latitude as originally demonstrated by Karasek (1979).
Schaubroeck and Merritt (1997)	Karasek's (1979) demand-control model Dependent variables: Diastolic blood pressure and systolic blood pressure as two distinct measures of job stress Independent variables: Job demands Job control Self-efficacy as moderating variable Decision latitude (freedom to choose	Multiple regression analysis using interaction terms Model 1: demographic variables Model 2: main effects Model 3: all two-way interactions Model 4: three-way interaction	110 randomly selected full-time health professionals at a large rehabilitation hospital	Main effects and two-way interaction effects of job demands, job control, and self-efficacy were nonsignificant in predicting diastolic and systolic blood pressures. The three-way interaction effect was, however, significant in predicting both diastolic ($p < .05$) blood pressure ($\beta = -14.6$, $\Delta R^2 = .07$, $F_{1,66} = 5.90$, $p < .018$, Total $R^2 = .24$) and systolic ($p < .01$) blood pressure ($\beta = -24.50$, $\Delta R^2 = .11$, $F_{1,66} = 9.68$, $p < .001$, Total $R^2 = .28$).

(*continued*)

Table 6.1 *(continued)*

Authors	Model/Variables Utilized	Methodological Approach	Sample	Findings
	method, opportunity to use ability, amount of variety, etc.)			
Schaubroeck and Fink (1998)	Karasek's (1979) demands-control model; Johnson and Hall's (1988) demands-control-support model <u>Dependent variables:</u> Job performance Prosocial organizational behavior (PSOB) (altruism, attendance, conscientiousness) Physical symptomatology Supervisor and intrinsic job satisfaction Organizational commitment Sick leave (hours) <u>Independent variables:</u> Social support (supervisor consideration); Employee control (control over quality, methods); Work (demands) stressors (quantitative workload, skill underutilization, role conflict, role ambiguity)	Hierarchical multiple regression analysis using interaction terms Model 1: main effects Model 2: 2-way interaction (support/control) Model 3: 3-way interaction (support/control/demands)	214 service counselors, sales counselors, and underwriters in two large insurance companies	The key results focused on significant three-way interactions. Consistent support was obtained for the notion that both job control and social support are needed to facilitate job demands and result in positive outcomes for organizational commitment ($F = 2.28$, $p < .06$) and satisfaction ($F = 3.79$, $p < .05$). Coping difficulties were reported for workers facing job demands coupled with high control and low support or low control and high support. Both control and support need to be high for effective coping. However, counter results are obtained for a single stressor—responsibility for others.
Singh (1998)	<u>Dependent variables:</u>	Path analysis using quadratic	518 sales and	Results varied by the job outcomes

	Job performance Job tension Job satisfaction Organizational commitment Turnover intentions <u>Independent variables:</u> Role conflict, role ambiguity, and role overload as three distinct measures of role stress Autonomy, participation, feedback, and task variety as four distinct measures of job characteristics	and interaction terms, and correcting for measurement error	marketing personnel drawn from the members of the Association of Sales and Marketing executives, and from two different units of a Fortune-500 industrial supplier firm	examined. Linear effects were supported for job satisfaction and organizational commitment. Two-way interactions involving role conflict and task variety, and role ambiguity and autonomy were supported for job performance ($p < .05$ and $p < .01$). Coping was more effective when high role conflict and high task variety, or high role ambiguity and low autonomy co-occurred. Similar two-way interactions, involving role overload/task variety and role ambiguity/task variety, were significant for turnover intentions ($p < .05$). Task variety facilitated coping with role overload but hindered it when role ambiguity was high. Curvilinear effects involving two-way interactions and quadratic terms were obtained with job tension as the dependent variable. Feedback facilitated coping with role conflict but hindered coping with role ambiguity ($p < .05$).
De Jonge, van Breukelen, Landeweerd, Nijhuis (1999)	Karasek's demands-control model <u>Dependent variables:</u> Emotional exhaustion Job-related anxiety Work motivation Job satisfaction	Multilevel regression analysis (VARCL) using interaction terms Model 1: Fully unconditional model only comprising of the random effects of units and institutions Model 2: All demographics and individual-level	1,489 randomly selected health care workers from different functional units (intensive care, surgical unit, etc.) from eight hospitals and eight nursing homes	In predicting emotional exhaustion and job-related anxiety, aggregate-level measures of job demands and job autonomy (included in Model 3) did not explain additional variance. In both cases, Model 2 had the highest predictive power ($R^2 = .25$, $R^2 = .09$, respectively), and only the main effect of job demands was significant ($p < .05$). The higher the demands, the higher were exhaustion and anxiety.

(*continued*)

Table 6.1 *(continued)*

Authors	Model/Variables Utilized	Methodological Approach	Sample	Findings
		measurements of job demands and autonomy		
	Independent variables:			In predicting work motivation and job satisfaction, aggregate-level measurements of job demands and job autonomy (included in Model 3) explained a significant amount of additional variance (R^2= .13 with ΔD = 11.71, $p < .05$, and R^2 = .13.2 with ΔD = 12.14, $p < .05$, respectively).
	Job demands	Model 3: Model 2 plus aggregate-level measurements of job demands and job autonomy		
	Job autonomy			
	Demands and autonomy are measured at both the individual and aggregate levels (mean group scores of different units included in the study). Aggregate-level measurements combine judgments across individual jobs, thus removing variance due to individual differences and idiosyncratic responses.	A likelihood ratio test (D) is used to determine the difference of predictive power across the models		The individual-level interaction effect of demands and autonomy and the aggregate-level main effect of demands significantly predicted motivation ($p <$ .05). The higher the interaction, the higher was motivation. The higher the aggregated demand, the lower was motivation.
				Individual-level main effects of job demand and job autonomy, and their aggregate-level interaction effect, significantly predicted job satisfaction ($p < .05$). The higher the individual-level autonomy and the higher the aggregate-level interaction, the higher was satisfaction. The higher the individual-level job demands the lower was satisfaction.
Janssen (2001)	Karasek's demands-control model	Hierarchical multiple	134 randomly	Job demands squared interacted with

	Dependent variables: Job performance (*standard* performance and *innovative* performance) Job satisfaction Both performance and satisfaction were measured by means of self-rated and leader-rated scales in addition to standard measures Independent variables: Job demands Effort-reward fairness perceptions	regression analysis using quadratic terms Interaction terms analysis is used to detect moderation and nonlinear effects	selected low-level and mid-level managers from a Dutch food producer	effort-reward fairness to predict both performance and satisfaction. The effect on standard job performance was significant ($p < .05$) with $R^2 = .30$. The effect on leader-rated innovative job performance was significant ($p < .05$) with $R^2 = .39$. For self-rated innovative job performance the effect was also significant ($p < .01$) with $R^2 = .25$. As for satisfaction, the effect was even more significant ($p < .001$) with an $R^2 = .28$. Post-hoc probing of these interaction effects indicated that the managers who perceive effort-reward fairness perform better and feel more satisfied in response to intermediate levels of job demands than managers who perceive "under-reward" unfairness.
Van Yperen and Hagedoorn (2003)	Karasek's demands-control model Dependent variables: Intrinsic motivation Fatigue Independent variables: Job demands Job control Social support	Hierarchical multiple regression analysis using interaction terms Interaction terms analysis is used to detect moderation effects	555 nurses who worked at specialized units for patients with mental deficiency	Job demand–job control interaction significantly predicted fatigue ($p < .001$) with $R^2 = .27$, but not intrinsic motivation. The demand-social support interaction, on the other hand, significantly predicted intrinsic motivation ($p < .01$), but not fatigue. Finally, the demand-control-support interaction predicted motivation (0 < .01), but not fatigue. Under conditions of low support, demands increased motivation when control was high, and decreased it when control was low. Under conditions of high support, demands decreased motivation when control was high. There was no effect when control was low.

Abdel-Halim (1981) found support for the interactive effects of role stressors (e.g., role conflict, ambiguity, overload) and job scope (e.g., skill variety, autonomy, feedback) on satisfaction across different technological contexts. Several other studies (De Jonge et al. 1999; Dwyer and Ganster 1991; Fox, Dwyer, and Ganster 1993) have supported Karasek's model by differentiating between the subjective and objective aspects of job demands and job control. In general, results are consistent with Dwyer and Ganster's (1991) findings that the interaction effects are consistently significant for the subjective or perceived measures of job demands and control (however, see De Jonge et al. 1999).

Other studies have provided evidence that either fails to support or refutes Karasek's contingency model. For instance, Spector (1987) studied the effects of demands and control on the satisfaction, frustration, anxiety, and health symptoms of clerical employees at a major university. He found that, out of thirty regressions, the interaction effect of demands and control was significant only in two, casting doubt on the validity of the demand–control model. Moreover, in these two cases of significant interaction effects, the direction of effects did not correspond to Karasek's hypothesis. Likewise, Fletcher and Jones (1993) were unable to find any significant interaction effect of demands and control on satisfaction, anxiety, and depression among patients at a health care center. Schaubroeck and Fink's (1998) study provided support for the three-way interactions of demands, control, and support, rather than the two-way interaction of demands and control as originally proposed by Karasek. Coping difficulties were found among employees facing high job demands coupled with high-control and low-support jobs. Likewise, low control coupled with high support produced coping difficulties in employees in high-demand jobs. As such, both control and support appear necessary for effective coping. In another study exploring three way effects involving demands, control, and support, Van Yperen and Hagedoorn (2003) reported that, while demand–control interaction influenced employee fatigue (but not motivation), the demand–support interaction had a significant effect on employees' intrinsic motivation (but not fatigue). In addition, a three-way interaction between demands, control, and support was significant for motivation but not fatigue. Depending on the level of job demands, control, and support, Van Yperen and Hagedoorn noted that some jobs or roles may evidence divergence between fatigue and motivation, such as when employees are highly motivated and fatigued at the same time.

Recently, researchers have modeled the curvilinear effects of job demands and job control on outcomes. For example, although he failed to find significant interaction effects of demands and control (e.g., decision latitude), Warr (1990) found that demands were nonlinearly (U-shaped) related to three separate dimensions of well-being among research employees—anxiety, depression, and displeasure. He also found that control was nonlinearly (increasing slope) related to job satisfaction. Curvilinear effects were more rigorously tested by Xie and Johns (1995), who found that job scope (e.g., task identity, task variety, autonomy, feedback)—a measure of job control—had a U-shaped relationship with job stress, measured by exhaustion and anxiety for a wide range of respondents (e.g., managers, sales workers, blue-collar workers). In other words, beyond an intermediate level, increasing levels of job scope enhance the stress level of boundary spanners instead of buffering the effects of role stressors as hypothesized by Karasek. These findings have been replicated and extended by Singh (1998).

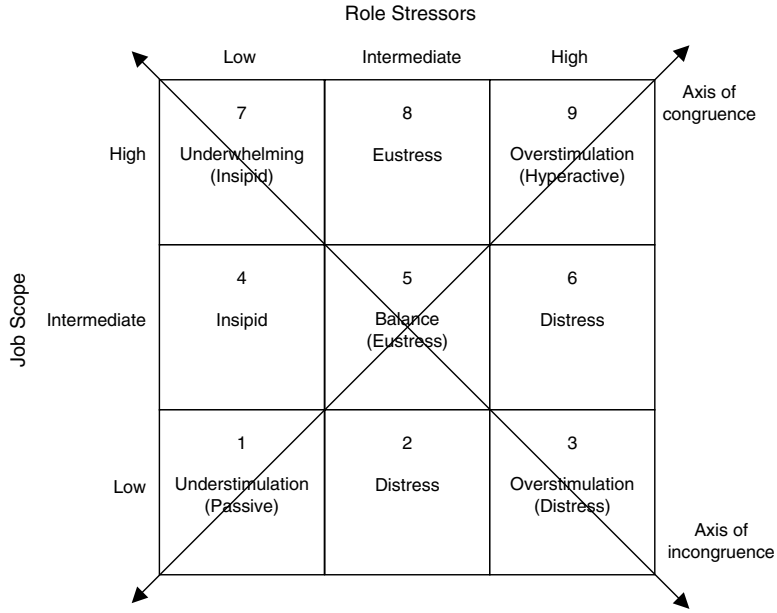
Assessment

Two themes emerge from our review of the empirical research rooted in Karasek's model. First, while the support for Karasek's model is mixed and uneven, this should be interpreted in light of the general lack of support for contingency models in organizational behavior (Tosi 1992). In many theories of organizational and human processes, interaction effects appear theoretically

plausible but fail to show up in empirical results. Researchers have indicated that this could occur for several reasons including (a) sampling, (b) measurement error, and (c) unnecessary complexity. In terms of sampling, the problem lies in the lack of heterogeneity in studies (e.g., samples from a single division/organization). Without the heterogeneity in perceived levels of job scope and role stressors, theoretical interaction effects are severely limited by the resulting restriction of range and fail to be empirically detected (Tosi 1992). In terms of measurement error, readers will note that most of the studies in Table 6.1 utilize regression analysis to examine interaction effects (Champoux 1992; Xie and Johns 1995). This is problematic because regression analysis ignores measurement error in dependent and independent variables. Because for most social science research, measurement error can be significant, ranging from 32 percent (Cote and Buckley 1988) to 50 percent (Schmidt and Hunter 1996), the regression coefficients are likely biased downward, making it difficult to empirically detect small but interesting contingency effects (Busemeyer and Jones 1983; MacCallum and Mar 1995). Finally, some researchers have argued that, despite the theoretical complexity of social science models, the empirical phenomena are governed by pragmatically simple and linear relationships. Because of random variation or unmodeled effects (e.g., contextual factors), it is plausible that some empirical studies could support interaction effects while many others refute such complexity resulting in a "mixed" pattern of results. In the context of Karasek's model, because researchers have not systematically tackled concerns due to sampling and measurement error, it is probably inappropriate to view the "mixed" evidence as support for the "unnecessary complexity" explanation at this time.

Second, empirical findings from the more recent studies suggest that Karasek's model needs to be modified to account for the curvilinear effects of job scope and role stressors on job outcomes. Both effects are supported by strong theoretical foundations. In the context of job scope, researchers draw from activation theory to posit an "overstimulation" effect so that excessive levels of job scope including feedback, participation, variety, and autonomy hinder rather help in one's performance (Champoux 1978, 1992; Schwab and Cummings 1976; Singh 1998). In turn, this overstimulation effect is based on three interrelated propositions: (1) job scope acts as a motivational force that stimulates an individual to increase effort or expend energy in task performance; (2) each individual has a "characteristic" level of stimulation that represents an optimal point of motivation; and (3) if experienced stimulation level substantially exceeds this "characteristic" level, the individual becomes overwhelmed, resulting in increased anxiety and reduced performance (Gardner and Cummings 1988; Kahn and Byosiere 1992). Consequently, both high (or excessive) and low (or inadequate) levels of job scope result in lowered performance and heightened anxiety (Xie and Johns 1995). Intermediate levels of job scope that are closer to the individual's characteristic level result in an optimal job context. Studies by Champoux (1992), Xie and Johns (1995), and others provide support for the curvilinear effects of job scope. In the case of role stressors, researchers evoke the Yerkes-Dodson law (Yerkes and Dodson 1908) drawn from early clinical and laboratory studies to posit that role stressors hold the potential for both "distress" and "eustress" (Selye 1976). According to this law, *both* at low and high levels of role stressors, an individual's job context is suboptimal. This is because performance is undermined by a lack of challenge in the "low" condition and by overactivation in the "high" condition. Moreover, both of the preceding conditions are characterized by passive coping driven by either a low level of motivation/resource activation or lack of sufficient resources to deal with overwhelming role stressors (Schaubroeck and Ganster 1993). This passive coping interferes with the individual's adaptivity to environmental demands, further deteriorating performance and eventually leading to a "distress" condition. However, performance is thought to be optimal in the *intermediate* role stress condition as the individual is energized/activated to respond and actively cope with environmental demands but is

Figure 6.2 **A Graphical Representation of the Contingency Model for the Influence of Role Stressors and Job Scope on Boundary Spanner Effectiveness**



not overwhelmed by them. Selye (1976) referred to this condition as "eustress" or "good" stress. Studies by Jamal (1984), Schaubroeck and Ganster (1993), and Singh (1998) provide support for the curvilinear effects of role stressors.

To simultaneously account for the curvilinear effects of role stressors and job scope, Karasek's model can be adapted by including an intermediate condition along both of its dimensions (see Figure 6.2). The resulting framework is a 3 x 3 model that represents nine distinct job contexts. For "low" levels of job scope, the job context is likely to be "passive," "distress," or "overwhelming" corresponding to "low," "intermediate," or "high" levels of role stressors (cells 1, 2, and 3). This is because a "low" job scope provides little to no coping resources to the boundary spanner and the distress level increases directly with role stressors, except in the event that role stressors are at a "low" level. In this instance, the job context lacks any stimulation whatsoever, resulting in a "passive" condition in accord with activation theory and Yerkes and Dodson's law. With "intermediate" level of job scope, the job context is either balanced (cell 5, scope = stressors) or unbalanced (cells 4 and 6, scope ≠ stressors). In the balanced condition, stressors and scope are at their optimal level and cohere with Karasek's notion of an "active" job context where stressors are high enough to provide the challenge without overwhelming the individual, and scope is high enough to aid coping with role demands without overstimulating the employee. This also accords with the inverted U-hypothesis. In the unbalanced conditions, the boundary spanner is either underwhelmed (scope > stressors) or overwhelmed (stressors > scope). In the former job context, role stressors are too low to challenge the individual, while in the latter condition, scope is probably insufficient to aid effective coping with role demands. Finally, under conditions of "high" scope, job contexts vary significantly depending on the level of role stressors ranging from a munificent and underwhelming job context (cell 7, scope > stressors) to a highly charged, overstimulated context where both scope and stressors are high (cell 9). Both are less than optimal as

per overstimulation and activation theories. However, for "intermediate" levels of role stressors, the job context is likely to indicate "eustress" as the job is both challenging and resourceful to allow effective coping with job demands (cell 8). Note that the model in Figure 6.2 is bisected by two diagonals: (a) an axis of congruence along which the job contexts are balanced with congruent levels of job scope and role stressors (i.e., cells 1, 5, and 9); and (b) an axis of incongruence depicting job contexts that are completely unbalanced with opposite levels of job scope and role stressors (i.e., cells 3 and 7). Interestingly, only a single optimal cell (cell 5) exists where both scope and stressors are at an intermediate level. While the preceding adaptation of Karasek's model is consistent with emerging empirical evidence and grounded in strong theory, it has not been subjected to empirical investigation.

Configural Perspective

We propose a configural perspective that also examines the effects of role stressors by explicitly considering the simultaneous and interactive influence of job scope. However, compared to the contingency perspective, a configural perspective is more complex because of several reasons. First, it accounts for unique patterns of role stressors and job scope (referred to as "configurations") that represent nonlinear and higher-order interactions that cannot be represented within the traditional contingency models (Miller and Friesen 1984). Second, the configurations are defined at a group rather than individual level such that they are shared representations of boundary roles. Third, it allows consideration of equifinality, or the notion that different configurations may be equally effective (e.g., in terms of performance, satisfaction). Fourth, a configural approach affords flexibility in modeling the phenomenon using approaches that distinguish between logical plausibility (e.g., ideal types) and empirical viability (Doty and Glick 1994). Below we discuss these ideas and review the limited empirical research to date.

Theory

In accord with configural theory, we posit that a job context can be defined by *any* specific combination of perceived role stressors (e.g., "high") and job scope (e.g., "moderate") that is a valid representation of boundary roles. Specifically, it posits that only a few, dominant combinations—termed "configurations"—are plausible that represent shared interpretive schemas of job contexts (Meyer, Tsui, and Hinings 1993). That is, a configurational perspective rejects the notion that infinite combinations of role stressors and job scope are *empirically* plausible as if these factors could be varied independently. Instead, it accepts the view that factors are interdependent and often can change only discretely. For instance, boundary roles involving extreme combinations of contrasts—such as high stress, low scope (cell 3, Figure 6.2) and high scope, low stress (cell 7, Figure 6.2)—are unlikely to be obtained empirically because they would be either so unbearable or unviable to be sustained as reasonable boundary roles in modern organizations. Likewise, it may be pragmatically difficult, if not impossible, to vary job conditions finely to obtain a continuous range of gradations. As we noted at the outset, practitioner views differ from academic thought on whether boundary roles can be redesigned as low-stress jobs. Practitioners' assertions that such redesign is difficult at best due to the stress inherent in boundary roles indicate empirical limits on the viable range of role stressors in boundary jobs. However, practitioners are not passive principals (e.g., managers) unconcerned about agent (e.g., boundary spanner) stress and effectiveness. Rather, as active mangers, practitioners actively design jobs to allow boundary spanners to cope with their role stressors and effectively serve organizational goals. Such design

efforts often involve supplying job scope to match role stressors, resulting in job contexts that gravitate toward the diagonal noted as the "axis of congruence" in Figure 6.2. Taken together, the preceding considerations suggest that only a few configurations of role stressors and job scope are likely to be empirically plausible.

Moreover, each configuration represents a schema about job context that is shared by collectives of boundary spanners. That is, a configuration is not defined at the individual level of analysis. Rather, it represents characteristics of the job that are shared by groups of boundary spanners. In this sense, configurations of job context are less sensitive to individual variability than operationalizations of job context in contingency or universalistic approaches. Debates about the appropriate level of analysis for conceptualizing job contexts have focused on two competing positions. On one hand, proponents of subjective experiences of work argue that an individual-level analysis is appropriate because it captures how boundary roles are perceived by individuals who fill them and any effort to aggregate experiences denies the fundamental place of individuals in organizations. On the other hand, job design researchers take a managerial perspective to argue for an organizational unit as the appropriate level of analysis to capture largely "objective" aspects of boundary roles populated in the unit and reject subjective experiences as noisy data that is less useful for managerial efforts in designing jobs. As the subjective-objective debate continues in the literature, the configural perspective offers an intermediate position that bridges these perspectives. The configural perspective does not deny the relevance of individual perceptions of boundary roles. The starting point for a configural perspective is the boundary role occupant's perceptions of role stressors and job scope inherent in his/her job. However, in construing job contexts as combinations of role stressors and job scope, the configural approach moves forward to identify configurations that are shared by collectives of boundary spanners. *Why would cohesive collective schemas emerge?* We posit that characteristics of boundary roles are interpreted by individuals through a process of interaction with job design practices, cognitive appraisals, and sense-making, and the resulting interpretations are shared, refined, and updated through a process of social interactions among boundary spanners resulting in shared collective schemas (James, Joyce, and Slocum 1988; Young and Parker 1999). Moreover, these collective schemas need not faithfully reproduce the objective work design features nor adhere to department/unit boundaries; yet, they meaningfully capture the patterns of boundary spanner interactions and systematically relate to critical employee and organizational outcomes (Young and Parker 1999). In this sense, configurations of boundary roles are jointly determined by objective work design efforts and subjective interpretations of boundary spanners.

Using a configural perspective to study job contexts has several advantages including testing for equifinality and modeling flexibility (Doty, Glick, and Huber 1993; Meyer, Tsui, and Hinings 1993; Miller and Friesen 1984). The notion of equifinality contrasts with the linear postulate that is common to most previous studies. The linear postulate posits that, given any two job contexts, it is possible to identify a single configuration that is ideally more favorable than the other because it provides more of the desirable characteristics (e.g., satisfaction, performance), and fewer of the undesirable characteristics (e.g., burnout, turnover intentions). In contrast, the equifinality proposition argues that it is highly unlikely that any configuration is significantly superior to other plausible configurations across all dependent variables considered. Rather, the more likely scenario is that two or more configurations are equally effective for some dependent variables (e.g., performance) and differentially effective for other variables (e.g., satisfaction). In other words, an equifinal view denies the presumed superiority of any specific configuration. In this sense, the proposition of equifinality embodies the notion of nonlinear effects. Such equifinality notions are, at best, difficult to test under universalistic and contingency approaches.

Finally, a configural perspective allows considerable flexibility in obtaining valid configurations including (a) deductive, theory-driven typology, and (b) inductive, empirical-driven taxonomy (Doty and Glick 1994). Typology-based configurations draw from a strong theoretical framework that specifies the number of plausible configurations, how each configuration is defined, and the rationale for the differential influence of different configurations. Because deductive configurations may or may not be empirically viable, they are often referred to as "ideal" types. By contrast, taxonomical configurations begin with data and extract viable configurations using aggregation procedures (e.g., cluster or latent class analysis). In order to obtain valid configurations, typical aggregation procedures are flexible to allow for (a) theory-based "centroids" to guide inductive procedures, and (b) split-sample designs that reduce the confounding effect of sampling variability. In this sense, not all logically plausible configurations can be empirically viable (Meyer, Tsui, and Hinings 1993). Readers will note that this concern with obtaining valid configurations when combined with tests for equifinality allows a finer-grained study of job contexts and their influence on boundary spanner processes than possible under universalistic or contingency perspectives.

Empirical Findings

Few, if any, studies have utilized a configural perspective to study role theory effects for boundary spanners. One such exceptional study is by Payne and Fletcher (1983). Using a sample of 148 teachers in UK, Payne and Fletcher utilized inductive procedures to extract empirically viable configurations. Foreshadowing later work, Payne and Fletcher utilized multiple dimensions of demands (e.g., disciplinary demands, maintaining standards, workload demands) and discretion (e.g., interpersonal support, job discretion) to faithfully capture the richness of teachers' job context. In all, seven distinct configurations of job contexts were obtained that differed in terms of the demands, discretion, and constraints. Unfortunately, Payne and Fletcher did not compare their taxonomical configurations with theoretically developed "ideal" job contexts. Nevertheless, they found that all of the obtained configurations were equifinal—that is, there were no differences in terms of outcomes, with a single exception. One configuration with low levels of demands and high levels of discretion produced a significantly *higher* level of job satisfaction than any other job context. Payne and Fletcher did not thoroughly investigate these counterintuitive results. In addition, while Payne and Fletcher pointed out the need for studies that sample heterogeneously from a well-defined sampling frame, they appear to be unaware of the empirical problems in detecting nonlinear and interaction effects with homogenous samples. No other study could be traced that had utilized a configural perspective to examine role theory effects.

Because the configural perspective approaches the phenomenon differently than contingency and universalistic approaches, it is likely that such a pursuit will yield new insights into the influence of role stressors and job scope in boundary spanning positions. Despite its shortcomings, the Payne and Fletcher study provides an initial indication that a configural perspective may be rewarding due to its potential to reveal equifinal effects. At the minimum, a configural perspective is likely to provide findings that would be useful to triangulate with contingency and universalistic findings to yield a holistic understanding of the stress processes among boundary spanners.

Potential Contributions and Concluding Notes

Looking at the body of literature on boundary role stress and its multiple meta-analyses, one is prone to conclude that room for potential contributions is limited and the hurdle steep. Our review is intended to dispel this view. Beyond the universalistic approach, vast areas remain

unexplored. Although the contingency perspective has received some attention, the study of contingencies that modulate the effect of role stressors for marketing-oriented boundary spanners has lagged. Marketing researchers appear to be more interested in identifying the different and emergent sources of role stress (e.g., work–family conflict) than in guiding boundary role design efforts that facilitate coping with role stressors or, more boldly, transforming role stressors into an energizing force for active and fulfilling engagement in boundary roles. Possibly, the terminology is a stumbling block. Socialized interpretations of stress appear to favor negative and avoidance representations that restrict construals of stress to dysfunctional mechanisms for boundary role effectiveness. Karasek's (1979) efforts to construe stress as role demands to blunt socialized representations appear to have had little impact in the marketing literature as evident from Table 6.1, where marketing studies are conspicuously absent. Alternatively, marketing researchers accept Karasek's theoretical arguments but are unmoved by its mixed empirical evidence. Lack of clear support for interaction effects does render the complexity of Karasek's contingency model unattractive. Complexity in theorizing with uncertainty in empirical payoffs poses another stumbling block.

Our review favors a view of the preceding stumbling blocks that suggests opportunities waiting to be exploited. Overriding socialized representation of stress that allows for its functional and eustress effects has considerable practitioner and theoretical appeal. To exploit this opportunity, marketing researchers might find it useful to coalesce around common terminologies that reframe the notion of stress in boundary-spanning positions. One option is to rid stress of its presumed negative or positive connotations, and to posit theoretically driven contingencies that result in its positive or negative effects. In this sense, whether stress is negative or positive is determined contextually subject to coping resources afforded by contexts and coping capabilities deployed by individuals operating in their boundary contexts.

Taking a totally different tack, another option is to develop alternative terminologies and redefine terms more precisely. A possible approach is to explicitly distinguish between *role stressors*—the degree to which boundary role characteristics pose demands on individual capabilities and resources, and require effortful coping, and *role stress*—the degree to which boundary spanners experience psychological and physiological symptoms indicating that role stressors are exceeding individual coping capabilities and resources (e.g., anxiety). As such, the concept of role stress retains its negative connotation and is indicative of job contexts that are dysfunctional for boundary role effectiveness because resources afforded by job contexts are not sufficient to facilitate boundary spanner coping. By contrast, role stressors can take on either a positive or negative meaning—as Karasek's original notion of role demands does—given the potential for distress or eustress depending on job scope or other contextual contingencies. We see promise in this approach and, indeed, have used it throughout this paper. We suggest its serious consideration in future studies to overcome terminological blocks.

Despite the mixed results of contingency hypotheses, several reasons suggest that this is a fruitful avenue for studying effectiveness of marketing-oriented boundary spanners. First, contingency effects are contextually dependent such that what works for organizational employees may or may not work for marketing-oriented boundary spanners. Customers represent a uniquely different boundary spanning problem compared to other boundary spanning roles involving working with either internal employees in other departments or suppliers. Customers introduce considerable heterogeneity and unpredictability to boundary exchanges. As such, boundary roles involving customers can present a unique profile of role stressors, and require considerable job scope to cope with inherent challenges. What is less salient in other boundary roles may well be quite prominent in marketing-oriented boundary spanners. Second, recent methodological advances have made it

feasible to empirically detect interaction effects by providing tractable approaches for controlling measurement error and testing nonlinear effects (Cortina, Chen, and Dunlap 2001; Ganzach 1997; Marsh, Wen, and Hau 2004). Current experience with these approaches suggests that disregarding measurement error severely hinders the detection of contingent effects, thereby providing biased results and rendering much of the previous research problematic. Third, more recent studies have reported support for the contingency hypotheses indicating that the influence of role stressors is not only contingent on job scope but also on other job factors including support (Van Yperen and Hagedoorn 2003). Together, the preceding developments point to potential payoffs from future studies that posit theory-based contingency hypotheses and examine them using methodological approaches that control for confounding influences including measurement error.

While the contingency approach requires reconsideration by marketing researchers, the configural approach represents a new frontier of unexplored possibilities. The configural approach begins from different assumptions about boundary role contexts and construes the study of boundary role effectiveness from a different lens, one that inherently accounts for nonlinear and higher-order interactions, allows collective-level perceptions of role dynamics, incorporates the notion of equifinality and affords flexibility in distinguishing between "ideal type" and "empirically viable" work contexts. We could not trace a single study in marketing that had utilized this approach for examining role theory effects for boundary spanners. Even in the extant body of work, empirical work on configural approach is lacking. We had traced a study by Payne and Fletcher (1983) that appeared consistent with this approach but neither explicitly nor fully considered the configural perspective. We view this gap as a significant opportunity for providing insights into role theory effects for marketing-oriented boundary spanners. To the extent that these insights compare and contrast with those obtained from universalistic and contingency perspectives, there opens a dialogue on triangulation efforts and on the ecological validity of assumptions underlying differing approaches. Exposing assumptions and pressing for a critical assessment of their empirical reasonableness holds considerable promise in advancing our understanding of role theory effects for marketing-oriented boundary spanners. Current studies locked in largely universalistic approach are incapable of raising such fundamental questions.

In closing, there is little disagreement about the burden and magnitude of individual, organizational, and societal cost of stress, especially on the boundaries that define an organization's interfaces with its customers and society at large. Disagreements arise both in identifying when role stressors exceed functional levels, and in selecting the "best" strategy for combating the consequences of boundary role stress. To bridge these disagreements and provide new insights, we provide a comparative review of different approaches, propose a configural approach that views job contexts as organizational or situational characteristics defined by specific stressor-scope combinations, and outline directions for future research that take advantage of the plurality of approaches. Overall, our review challenges marketing researchers to explicitly consider *how boundary spanners cope with role stressors* when they aim to examine *what are the consequences of role stressors for boundary spanners.* Examining role stressors without considering boundary spanner coping efforts is like studying water level in a bucket by observing the inflows without considering the outflows. Coping with role stressors is effortful and demands resources that either come from job scope supplied by the context or from an internal reservoir, which results in depletion afterward. It can be argued that organizations are responsible for providing a balanced job context where job scope provides sufficient resources to cope with role stressors without depleting internal reservoirs of boundary role occupants. Toward this end, our review implores marketing researchers to adopt a role design perspective that considers how job context can be configured for optimal effectiveness. Boundary role stress is not an individual problem, nor is it just a matter to be addressed organizational

design. Rather, it requires a simultaneous consideration of job context and individual factors to develop active, challenging, and resourceful jobs that fulfill Karasek's ambition for healthy work. We hope our study provides the impetus for such pursuit.

References

Abdel-Halim, Ahmed A. 1981. "Effects of Role Stress-Job Design-Technology Interaction on Employee Work Satisfaction." *Academy of Management Journal* 24: 260–273.

Behrman, Douglas N., and William D. Perreault. 1984. "A Role Stress Model of the Performance and Satisfaction of Industrial Salespersons." *Journal of Marketing* 48: 9–21.

Belasco, James E. 1966. "The Salesman's Role Revisited." *Journal of Marketing* 30: 6–11.

Blood, Milton R., and Charles L. Hulin. 1967. "Alienation, Environmental Characteristics, and Workers' Responses." *Journal of Applied Psychology* 51: 284–306.

Brown, Steven P., and Robert A. Peterson. 1993. "Antecedents and Consequences of Salesperson Job Satisfaction: Meta-Analysis and Assessment of Causal Effects." *Journal of Marketing Research* 30: 63–77.

Busemeyer, Jerome R., and Lawrence E. Jones. 1983. "Analysis of Multiplicative Combination Rules When the Casual Variables Are Measures with Error." *Psychological Bulletin* 93: 549–562.

Cavanaugh, M.A., W.R. Boswell, M.V. Roehling, and J.W. Boudreau. 2000. "An Empirical Examination of Self-Reported Work Stress among U.S. Managers." *Journal of Applied Psychology* 85: 65–74.

Champoux, Joseph E. 1978. "A Preliminary Examination of Some Complex Job Scope-Growth Need Strength Interactions." *Academy of Management Proceedings* 38: 59–63.

———. 1992. "A Multivariate Analysis of Curvilinear Relationships Among Scope Work Context Satisfaction and Affective Outcomes." *Human Relations* 45: 87–111.

Churchill, Gilbert A., Neill M. Ford, Steven W. Hartley, Orville C. Walker. 1985. "Determinants of Sales Performance." *Journal of Marketing Research* 22: 103–118.

Cortina, Jose M., Gilad Chen, and William P. Dunlap. 2001. "Testing Interaction Effects in LISREL: Examination and Illustration of Available Procedures." *Organizational Research Methods* 4: 324–360.

Cote, Joseph A., and M. Ronald Buckley. 1988. "Measurement Error and Theory Testing in Consumer Research: An Illustration of the Importance of Construct Validation." *Journal of Consumer Research* 14: 579–582.

Csikszentmihalyi, Mihaly. 1975. *Beyond Boredom and Anxiety.* San Francisco, CA: Jossey-Bass.

Daniels, Kevin, and Andrew Guppy. 1994. "Occupational Distress, Social Support, Job Control, and Psychological Well-Being." *Human Relations* 47: 1523–1544.

De Jonge, Jan, Gerard J.P. van Breukelen, Jan A. Landeweerd, and Frans J.N. Nijhuis. 1999. "Comparing Group and Individual Level Assessments of Job Demand-Control Model: A Multilevel Approach." *Human Relations* 52: 95–122.

Doty, D. Harold, and William H. Glick. 1994. "Typologies as a Unique Form of Theory Building: Toward Improved Understanding and Modeling." *Academy of Management Review* 19: 230–251.

Doty, D. Harold, William H. Glick, and George P. Huber. 1993. "Fit, Equifinality, and Organizational Effectiveness: A Test." *Academy of Management Journal* 36: 1196–1250.

Dwyer, Deborah J., and Daniel C. Ganster. 1991. "The Effects of Job Demands and Control on Employee Attendance and Satisfaction." *Journal of Organizational Behavior* 12: 595–608.

Fisher, Cynthia D., and Richard J. Gitelson. 1983. "A Meta-Analysis of the Correlates of Role Conflict and Ambiguity." *Journal of Applied Psychology* 68: 320–333.

Fletcher, Ben C., and Fiona Jones. 1993. "A Refutation of Karasek's Demand-Discretion Model of Occupational Stress with a Range of Development Measures." *Journal of Organizational Behavior* 14: 319–330.

Fox, Marilyn L., Deborah J. Dwyer, and Daniel C. Ganster. 1993. "Effects of Stressful Job Demands and Control on Psychological and Attitudinal Outcomes in a Hospital Setting." *Academy of Management Journal* 36: 289–312.

Ganzach, Y. 1997. "Misleading Interaction and Curvilinear Terms." *Psychological Methods* 3: 235–247.

Gardner, Donald G., and Larry L. Cummings. 1988. "Activation Theory and Job Design: Review and Reconceptualization." In *Research in Organizational Behavior*, ed. B.M. Staw and L.L Cummings, 81–122. Greenwich, CT: JAI Press.

Hackman, J. Richard, and Greg R. Oldham. 1976. "Motivation through the Design of Work: Test of a Theory." *Journal of Applied Psychology* 59: 638–642.

Hobfoll, Stevan E., and John Freedy. 1993. "Conservation of Resources: A General Stress Theory Applied to Burnout." In *Professional Burnout: Recent Developments in Theory and Research*, ed. Christina Maslach, Wilmar B. Schaufeli, and Tadeusz Marek, 115–133. Washington, DC: Taylor & Francis.

Jackson, Susan, and Randall Schuler. 1985. "A Meta-Analysis and Conceptual Critique of Research on Role Ambiguity and Role Conflict in Work Setting." *Organizational Behavior and Human Decision Process* 36: 16–78.

Jamal, Muhammad. 1984. "Job Stress and Job Performance Controversy: An Empirical Assessment." *Organizational Behavior and Human Decision Processes* 33: 1–21.

James, Lawrence R., William F. Joyce, and John W. Slocum. 1989. "Comment: Organizations Do Not Cognize." *Academy of Management Review* 13: 129–132.

Kahn, Robert L., and Philippe Byosiere. 1992. "Stress in Organizations." In *Handbook of Industrial and Organizational Psychology*, ed. Marvin D. Dunnette and Leaetta M. Hough. Palo Alto, CA: Consulting Psychologists Press.

Kahn, Robert L., Don Wolfe, Robert Quinn, and J. Diedrick Snoek. 1964. *Organizational Stress: Studies in Role Conflict and Ambiguity.* New York: John Wiley & Sons.

Karasek, Robert A. 1979. "Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign." *Administrative Science Quarterly* 24: 285–308.

Kelloway, E. Kevin, and Julian Barling. 1991. "Job Characteristics, Role Stress and Mental Health." *Journal of Occupational Psychology* 64: 291–304.

King, L.A., and D.W. King. 1990. "Role Conflict and Role Ambiguity: A Critical Assessment of Construct Validity." *Psychological Bulletin* 107: 48–64.

Landsbergis, Peter A. 1988. "Occupational Stress Among Health Care Workers: A Test of the Job Demands-Control Model." *Journal of Organizational Behavior* 9: 217–239.

MacCallum, Robert C., and Corinne M. Mar. 1995. "Distinguishing Between Moderator and Quadratic Effects in Multiple Regression." *Psychological Bulletin* 118: 405–421.

Marino, S. 1997. "The Stress Epidemic." *Industry Week* 246: 14.

Marsh, Herbert W., Zhonglin Wen, and Kit-Tai Hau. 2004. "Structural Equation Models of Latent Interactions: Evaluation of Alternative Estimation Strategies and Indicator Construction." *Psychological Methods* 9: 275–396.

Maslach, Christina, and Michael P. Leiter. 1997. *The Truth About Burnout.* San Francisco, CA: Jossey-Bass.

Merton, Robert K. 1949. *Social Theory and Social Structure: Toward the Codification of Theory and Research.* Glencoe, IL: Free Press.

Meyer, Alan D., Anne S. Tsui, and C. Roy Hinings. 1993. "Configurational Approaches to Organizational Analysis." *Academy of Management Journal* 36: 1175–1195.

Miller, Danny, and Peter H. Friesen. 1984. *Organizations: A Quantum View.* Englewood Cliffs, NJ: Prentice-Hall.

Mohrman, Susan A., and Susan G. Cohen. 1995. "When People Get Out of the Box: New Relationships, New Systems." In *The Changing Nature of Work*, ed. A. Howard. San Francisco, CA: Jossey-Bass.

Newton, Tim. 1995. *Managing Stress: Emotion and Power at Work.* Thousand Oaks, CA: Sage.

Payne, Roy, and Ben C. Fletcher. 1983. "Job Demands, Support, and Constraints as Predictors of Psychological Strain Among Schoolteachers." *Journal of Vocational Behavior* 22: 136–147.

Rhoads, Gary K., Jagdip Singh, and Phillips Goodell. 1994. "The Multiple Dimensions of Role Ambiguity and Their Impact on Psychological and Behavioral Outcomes Industrial Salespeople." *Journal of Personal Selling & Sales Management* 14: 1–24.

Rizzo, John R., Robert House, and Stephen I. Lirtzman. 1970. "Role Conflict and Ambiguity in Complex Organizations." *Administrative Science Quarterly* 15: 150–163.

Rommetveit, Ragnar. 1954. *Social Norms and Roles.* Minneapolis: University of Minnesota Press.

Schaubroeck, John, and Laurence S. Fink. 1998. "Facilitating and Inhibiting Effects of Job Control and Social Support on Stress Outcomes and Role Behavior: A Contingency Model." *Journal of Organizational Behavior* 19: 167–195.

Schaubroeck, John, and Daniel C. Ganster. 1993. "Chronic Demands and Responsivity to Challenge." *Journal of Applied Psychology* 78: 73–85.

Schmidt, Frank L., and John E. Hunter. 1996. "Measurement Error in Psychological Research: Lessons from 26 Research Scenarios." *Psychological Methods* 1: 199–223.

Schwab, Donald P., and Larry L. Cummings. 1976. "A Theoretical Analysis of the Impact of Task Scope on Employee Performance." *Academy of Management Review* 1: 23–35.

Selye, Hans. 1976. *The Stress of Life.* New York: McGraw-Hill.

Singh, Jagdip. 1993. "Boundary Role Ambiguity: Facets, Determinants and Impacts." *Journal of Marketing* 57: 11–31

———. 1998. "Striking a Balance in Boundary-Spanning Positions: An Investigation of Some Unconventional Influences of Role Stressors and Job Characteristics on Job Outcomes of Salespeople." *Journal of Marketing* 62: 69–87.

Singh, Jagdip, Jerry R. Goolsby, and Gary K. Rhoads. 1994. "Behavioral and Psychological Consequences of Boundary Spanning Burnout for Customer Service Representatives." *Journal of Marketing Research* 31: 558–569

Spector, Paul E. 1987. "Interactive Effects of Perceived Control and Job Stressors on Affective Reactions and Health Outcomes for Clerical Workers." *Work and Stress* 1: 155–162.

Spector, Paul E., Deborah J. Dwyer, and Steve M. Jex. 1988. "Relation of Job Stressors to Affective, Health, and Performance Outcomes: A Comparison of Multiple Data Sources." *Journal of Applied Psychology* 73: 11–19.

Tosi, Henry. 1992. *The Environment/Organization/Person Contingency Model: A Meso Approach to the Study of Organizations.* Greenwich, CT: JAI Press.

Tubre, Travis C., and Judith M. Collins. 2000. "Jackson and Schuler (1985) Revisited: A Meta-Analysis of the Relationships Between Role Ambiguity, Role Conflict and Job Performance." *Journal of Management* 26: 155–169.

Van Yperen, Nico W., and Mariet Hagedoorn. 2003. "Do High Job Demands Increase Motivation or Fatigue or Both? The Role of Job Control and Job Social Support." *Academy of Management Journal* 46: 339–348.

Warr, Peter B. 1990. "Decision Latitude, Job Demands, and Employee Well-Being." *Work and Stress* 4: 285–294.

Xie, Jia L., and Gary Johns. 1995. "Job Scope and Stress: Can Job Scope Be Too High?" *Academy of Management Journal* 38: 1288–1309.

Yerkes, Robert M., and John Dodson. 1908. "The Relation of Strength of Stimulus to Rapidity of Habit Formation." *Journal of Comparative Neurology and Psychology* 10: 459–482.

Young, Scott A., and Christopher P. Parker. 1999. "Predicting Collective Climates: Assessing the Role of Shared Work Values, Needs, Employee Interaction, and Work Group Membership." *Journal of Organizational Behavior* 20: 1199–1218.

CHAPTER 7

DESIGNING PRICE CONTRACTS FOR PROCUREMENT AND MARKETING OF INDUSTRIAL EQUIPMENT

GEORGE JOHN

Abstract

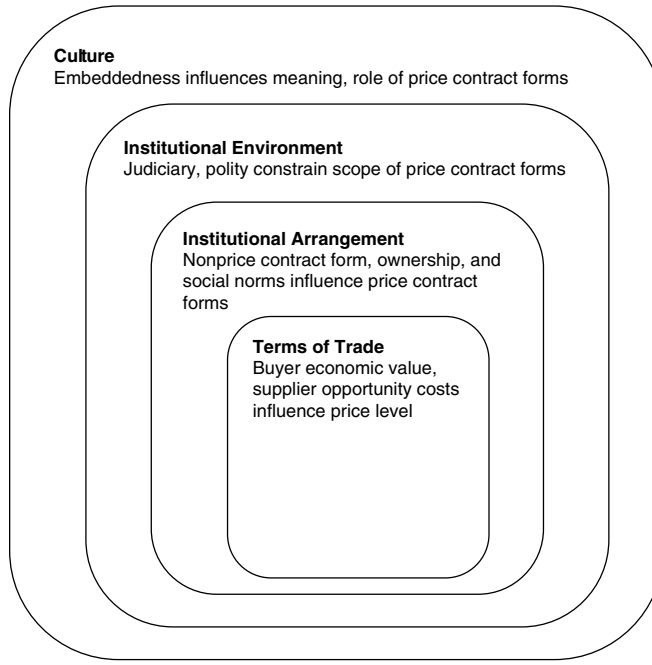
A durable goods contract typically specifies a price determination process, which is an integral part of "rules of the game" that govern the exchange between the contracting parties. Drawing on principles of new institutional economics, these contractual price processes are arrayed on a flexibility continuum ranging from predetermination through redetermination to renegotiation. A well-designed contract specifies a price process that is aligned in a discriminating fashion with the two fundamental problems of efficient exchange—adapting to changed circumstances, while simultaneously safeguarding against opportunism. A simple but effective price design rule is as follows: Greater adaptation needs call for moving away from predetermination processes toward renegotiation processes, while increased opportunism hazards call for moving in the opposite direction. The proper balance is illustrated with two cases involving durable goods. The choice between cost-plus price contracts versus fixed price contracts is shown to yield to this design principle as does the choice between outright sales contracts versus operating leases contracts.

Introduction

Modern marketing textbooks have long neglected "what" questions of pricing in favor of attending to the more readily understood and communicated "how much" question. To illustrate, one can readily find schemes to compute the optimal spending on media advertising, but much less attention paid to the corresponding question of whether to employ media advertising or to utilize an alternative such as on-site sales efforts. This is a natural outgrowth of the development in marginal analysis techniques. Ironically, the early research in marketing emphasized institutional matters (e.g., Breyer 1934), but these works were largely unable to provide empirically testable propositions and thus did not sustain a research enterprise. Much of this has changed in the last two decades due to the success of newer institutional[1] theories. For instance, contemporary marketing textbooks (e.g., Coughlan et al. 2001) cite transaction cost to explain vertical integration patterns in channels of distribution. However, the pricing domain of problems in marketing has resisted this resurgence of the use of institutional analysis.

Textbook orthodoxy as well as working paper orthodoxy on pricing issues hews to the marginal analysis tradition of neoclassical economics. The dictum of selecting prices that equate marginal revenues and marginal cost is central to this work. For the most part, contemporary efforts by marketing scholars continue in this tradition, albeit under more realistic, complex assumptions made

Figure 7.1 **Effects of Institutional Levels on Price Contract Form**



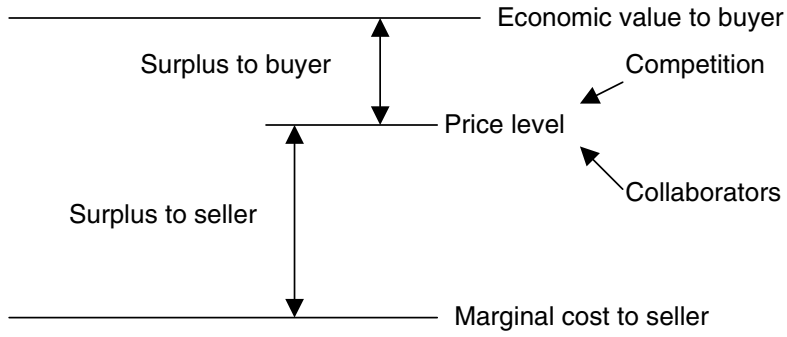
about consumers and firms. Thus, reference-dependent prices, prices with information asymmetries, and pricing involving strategic actors are commonly explored in contemporary work. However, the level of prices is still the outcome sought from these investigations. Unlike the specification of a particular number for a price level, the price format decision is the specification of the rules under which the price itself will be computed or determined. These are most germane to negotiated contracts between individual buyers and sellers. For instance, a canonical choice within negotiated contracts is the choice of a fixed price versus a cost-plus price. This format question has attracted very little attention in marketing, and contemporary textbooks offer little in the way of usable frameworks for unpacking the pros and cons of different price formats.

This chapter begins with the argument that price format choices precede the selection of a price level. These price formats are an integral aspect of the institutional arrangement devised to govern an exchange. Following a brief review of institutions (particularly their pricing elements), we articulate a design protocol, which is then illustrated by applying it to (a) the choice of fixed versus cost-plus prices for procuring components from a supplier, and (b) the choice of leasing versus selling price formats for industrial equipment.

Pricing Elements of Institutions

Institutions are defined as the rules that govern the interaction between economic actors. There are multiple levels of such rules that surround an exchange. Williamson (1999) describes the four concentric layers of institutions in Figure 7.1. These levels are arranged in order of their temporal horizon, with the outermost level changing at the slowest speed. Thus, each level is increasingly resistant to change within a given exchange. Put differently, each level is an exogenous constraint

Figure 7.2 **Value-Cost Framework for Price Decisions**



on the level below it. There are multiple aspects of these rules at each level, including formal and informal elements. Below, we sketch the price aspects of institutions at each of these nested levels. We lean heavily on Carson et al. (1999) in this section.

Price Elements of Terms of Trade

At the innermost level, exchange-specific rules define the *terms of trade* (specific prices, quantity and quality levels of the traded items) of the exchange. Neoclassical economic analysis insists that choosing the "right" price (e.g., choosing prices that equate marginal revenue with marginal cost) is the principal economic problem at this level; furthermore, applying this principle this will further economic well-being. The right price plays two key roles at this level. It divides the economic surplus (gains from trade) between the trading partners and it motivates a continuing supply of the good. Such insights have been embedded into managerial protocols for making price decisions in marketing. Consider, for instance, the framework in Figure 7.2, which is derived from the textbook by Nagle and Holden (2002).

Figure 7.2 is a valuable decision framework because practical methods have been developed to compute constructs such as the value to the buyer (e.g., using economic value-in-use analysis, willingness-to-pay conjoint, etc.), and the marginal costs of the seller (e.g., using activity-based cost analysis, etc.).

The most underdeveloped aspect of Figure 7.2 is the lack of any explicit consideration of the "what" (price format) issue. Implicitly, it assumes that we seek a selling price level *given* some format decision. As such, alternative price formats such as fixed versus cost-plus prices, unit prices versus all-you-can-eat/subscription prices, or rental versus selling plans are all assumed away. Notice that in each instance, these formats are assumed to have been chosen prior to choosing the level of price within the chosen format. Thus, we need to unpack the constraints imposed by these prior choices emanating from the institutional arrangement. We note that these aspects of price are sometimes termed "price metrics" (e.g., Nagle and Holden 2002), but there is no systematic work into their determinants or effects.[2]

Price Elements of Institutional Arrangements

The second level of Figure 7.1 describes the *institutional arrangement* devised by the parties to govern their exchange. This consists of the exchange-specific *contracting* rules, *ownership*

(property rights) rules, and *social* rules within which the terms of trade described above are embedded. Prices are a key aspect of each of these categories of rules, as we shall elaborate in this section.

Posted Prices, Auction Prices, and Contract Prices

There are three macro or broad types of price formats found in practice (Milgrom 1989): posted prices, auctions, and contract prices. Posted prices are take-it-or-leave-it prices announced to all comers, while auctions reveal the price after the event. Contracts are specifically devised for a particular exchange as the outcome of some negotiation or bargaining process.[3] Each of these types exhibits variation as well.

Posted price formats include linear price formats, quantity-based formats, subscription formats (e.g., post-paid cell phone plans), and fixed fee (all-you-can-eat) prices. Much of this complexity in posted prices is still amenable to the traditional marginal calculus approach, so we shall not focus on these choices, but we will note that these different price metrics do not involve a change in the price format (viz. posted prices or contracts) itself.

Auctions also exhibit a wide variety of complexity in their pricing elements as well (e.g., first price versus second price; open-outcry versus sealed-bid). Despite the growing popularity of auctions, it is a highly specialized topic, and we will not focus on it in this paper because auctions are never implemented between a single buyer and single seller, which is the focus of our effort. The institutional arrangement that is devised by one buyer and one seller to govern their exchange is a negotiated contract.

Negotiated contracts govern price determination rules between individual buyers and sellers, particularly in business-to-business exchanges. The delivered price is determined by the application of the price determination rules embedded within the terms of the negotiated contract in effect at the time of shipment. There is an enormous variety of contract price formats found in practice.

Appendix 7.1 illustrates a small part of one contract for natural gas between a well owner (seller) and a pipeline (buyer) in which the price determination rules are explained. It is crucial for us to order these contract alternatives in a parsimonious way. We borrow from the work on relational contracting to accomplish this task. Predetermination, redetermination, and renegotiation formats are three basic categories of price contract forms that can be arranged along a continuum.

Continuum of Negotiated Contract Formats

In a very influential study, Crocker and Masten (1991) showed that much of the complexity in price contract formats can be captured by locating them along a dimension of flexibility/incompleteness. The location of a particular contract form on this continuum *defines the degree to which prices for future transactions are known with certainty at the outset.*

At one end of this continuum are contract forms that are very rigid, and that *predetermine* prices over the contract horizon from the outset. A firm, fixed price that holds for the duration of the contract anchors this end, and constitutes the prototype of a predetermination price rule.

As we move along the continuum, one finds formulaic adjustment formats that allow prices to change over the contract horizon, but nevertheless employ formulae that are known ex ante. Escalator clauses, inflation adjustments, and other indexed price formats are examples of these *redetermination* price rules.

Further along the incompleteness continuum, we find price contract rules that do not prede-

termine nor redetermine price levels according to formulae that are devised at the outset. Rather, we find contracts that specify the *renegotiation* process. Renegotiations constrained by formulaic ceilings and/or floors and veto clauses are examples of these renegotiation price rules.

At the other end of the continuum, we find that the renegotiation processes are very open-ended and express generic constraints along the lines of good faith and reasonableness. Anchoring this end of the continuum are cost-plus price contract forms.

This categorization scheme successfully orders price contract formats reported in a number of recent empirical studies. Rosen (2004) found that contracts governing early-stage ties between original equipment manufacturers (OEMs) and suppliers embodied flexible (redetermination and/or renegotiation) price terms and fixed (predetermined) price terms in roughly equal proportions. Crocker and Reynolds (1993) show that the government employed a cost-plus price format in the round of initial contracts for a new generation of jet fighter engines from Pratt-Whitney, but that fixed price formats were employed for later contract rounds. Finally, Ghosh (2006) found that sellers of knowledge and service-intensive industrial equipment employed a mixture of lease price formats and outright sales price formats in roughly equal proportions. In all these instances, the parties appear to have selected their contracts in response to the circumstances surrounding the transaction.

Choosing a Contract Format

The mechanism design literature in economics has long emphasized information asymmetry and incentive alignment as the drivers of price contract forms. This literature provides us considerable theoretical insights into the manner in which contract pricing can be fine-tuned so as to mitigate problems arising from the specific structure of hidden information and hidden action in a given context (e.g., Laffont and Tirole 1993). However, the empirical evidence points to a much coarser selection of contract terms. Inspection of real-world contracts (e.g., Bajari, McMillan, and Tadelis 2002) rarely reveals the kind of finely tuned incentive schemes that are derived from mechanism design models. For instance, Walden (2002) found that large outsourcing contracts rarely included incentive clauses.

Bajari and Tadelis (2001) argue that the adaptation properties of negotiated contracts are probably more important than the incentive and screening properties of contract terms. It is the assumption of complete contracting[4] that drives much of the fine-grained incentive contracting schemes in mechanism design models, but in the real world, the object to be exchanged or project to be undertaken is often only partially or incompletely specified and/or is subject to revision during the execution stage. As such, the incentive effects of contracts are muted because of renegotiation issues (e.g., Williamson 1999) that are injected into the setting. The available empirical evidence suggests that this adaptation lens fits the empirical data much better.

In this latter view, alternative price contract forms are chosen primarily because of their differential ability to enable adaptations during contract execution. Using a formal model, Bajari and Tadelis (2001) show that it is easier to change the price under a cost-plus scheme than under a fixed price scheme because the parties have to track only the sum of the costs of the items that are eventually produced (including missteps) versus backing out the price of missteps out of the total sum of costs as would be required under a fixed price scheme.

The core insight from this literature boils down to a trade-off between the superior incentives (for cost reduction, etc.) provided by relatively fixed/formulaic price terms versus the superior adaptation properties of flexible/renegotiable price terms. We shall use this insight as a central principle in our decision framework for choosing price contract forms.

Ownership Aspects of Institutional Arrangements

An institutional arrangement does not only rely on price determination rules to govern interaction. Another vital aspect is the ownership of assets, including those tangible and intangible assets used for production as well as the traded asset itself.

At a basic level, ownership matters because contracts are always incomplete. In discussing ownership of production assets from an institutional standpoint, Carson et al. (1999) note that ownership is far more than a claim to residual profits. Rather, it permits the owner to direct (and redirect) the use of a productive asset, which then motivates greater investment and effort with respect to that asset as documented in the property rights branch (e.g., Grossman and Hart 1986) of the institutional literature.

Our interest is not in the ownership of the production assets, but rather in the ownership pattern of the traded item. In the classic exchange paradigm, the ownership of an item passes from buyer to seller in return for the price paid for the item. However, this masks the variety of patterns found in practice. In particular, we find alternative price rules that turn on ownership patterns. These patterns can be placed along a continuum as per Crocker and Masten (1991).

Continuum of Ownership Formats

The seller initially owns an asset that is to be transferred to a buyer for a price. The degree to which ownership rights are transferred to the buyer locates the particular contract form on the continuum. Consider the alternative forms commonly found in the case of durable equipment.

At one end of the continuum, we find a sale format. Here, the control of the asset passes completely to the buyer, who may then choose to use, maintain, and resell the asset as they see fit. Restrictions on one or more of these rights may accompany the exchange, which then moves the contract form along the continuum.

One discernable shift occurs when the exchange is structured as a lease. In particular, consider an operational lease (as distinct from a financial lease[5]). In an operational lease, the legal ownership of the asset remains with the seller, while the buyer is obligated to make periodic payments for the use of the asset. The form of these payments may vary considerably, but they typically possess the right to terminate the payments and return the asset (perhaps with some cancellation penalties).

At the other end of the continuum, one finds that the constellation of usage rights that pass to the buyer are quite limited in duration and scope even in comparison to operational leases. Most short-term rental contracts fit this description. Likewise, in intangible goods, the restrictive licensing schemes also fit this end of the continuum.

Choosing an Ownership Format

In a neoclassical world, transfer rights are simply another attribute of the product that is priced. In other words, there is some price at which each particular right would transfer. Thus, demand for goods conveying some set of rights with it would be different from demand for the same good with a different set of rights. However, recent work has emphasized that different formats can vary in their fit with particular circumstances. The mechanism design literature has emphasized the use of leasing as turning on the desire of the seller to price-discriminate among its customers by solving the so-called commitment or Coasian problem (e.g., Waldman 2003). However, as Ghosh (2006) and Masten and Snyder (1993) show, the empirical patterns are much more supportive of an institutional explanation for these leasing practices. Here, it is the superior adaptation proper-

ties of operational leases over sales formats that drive their use. Thus, when these properties are important enough to overcome other drawbacks of leases (e.g., assets being run down by lessees), they are favored over outright sales.

Price Elements of Institutional Environments

At the next level of institutions in Figure 7.1, one finds the institutional environment, which is defined by the structure of the judiciary and the polity of the nation-state within which the transaction occurs. These structures constrain the form of the institutional arrangement devised at the immediately lower level.

Briefly, the independence of the judiciary strengthens and thus expands the scope of contract forms that can be employed. Likewise, multipolar (federal) polities strengthen the predictability of contract enforcement, which also expands the scope of contract forms that can be employed. There is a growing body of evidence that the institutional environment has a particularly significant influence on price formats, contract forms, and outcomes for transactions that are quasi-public and/or that involve the government as one actor. Regulated industries, infrastructure, and public welfare initiatives are all instances of such transactions.

Levy and Spiller (1994) show that countries with independent judiciaries attract much greater private investment into their telecommunications industry, which then greatly improves the quality of the product. Indeed, at the macro level, the dominant view (e.g., North 1991) is that the political and judicial structures are the most significant determinants of a nation's economic development. Unfortunately, these institutions are difficult to engineer, so they are best viewed as constraints to a price contract designer.

Price Elements of Culture

At the outermost level of Figure 7.1, we find the cultural or societal norms that support exchange. Here, too, there is growing evidence that trust and related norms of exchange matter greatly to the economic well-being of societies. Studies from economic anthropology (e.g., Ellickson 1991) provide compelling evidence that societies with cooperative exchange norms improve trade dramatically. However, these norms are very difficult to engineer explicitly, thus they remain outside the grasp of the contract designer. Nevertheless, they remain formidable constraints on the design of institutional arrangements.

Parenthetically, it is important to observe the conceptual difference between the exogenous societal level norms and exchange-specific social rules. The former speak to the extant level of supportive social norms that are present in the society at large, whereas the latter speak to the actual social norms that are activated in a specific exchange. It is possible to activate or engineer cooperative norms from the larger culture through a conscious design of the institutional arrangement. For instance, Bercovitz, Jap, and Nickerson (2006) show that the engineering of symmetric investments by both parties to an exchange reliably evokes cooperative norms.

Summary

The main action for an institutional designer rests at the level of the institutional arrangement. The institutional environment and the cultural level above the institutional arrangement are largely exogenous to an exchange, so we need to accommodate their constraints on our choice of price contract terms. At the level below the contract form, the marginal analysis paradigm informs the

price decision quite well. Thus, we do not speak directly to the actual level of prices below. Rather, we seek to develop design protocols for designing price contract forms for specific marketing transactions. Two such design choices are considered; however, we turn first to the task of describing briefly the underlying common principles of governance structure design.[6]

Governance Design Templates

Economic efficiency is the primary impulse that drives the configuration of institutional arrangements. Operationally, the design criterion is the minimization of the transaction costs of exchange, including both ex ante costs as well as ex post costs.

Ex ante transactions costs consist of the "ink" costs of reaching enforceable agreements and the opportunity costs of insufficient investments made by actors whose interests are not sufficiently safeguarded. Ex post transaction costs consist of enforcement costs related to shirking, the bargaining costs of revising previous agreements, and the opportunity costs of not undertaking profitable revisions. These costs are minimized by matching up governance structures with the attributes of the exchange using the following matching rules or design principles.

1. Governance structures must protect the exchange-specific investments required for least cost production.
2. Governance structures must match the need for revising initial designs and plans during the execution stage.
3. Governance structures must match the coarseness of the verifiable performance metrics at hand.

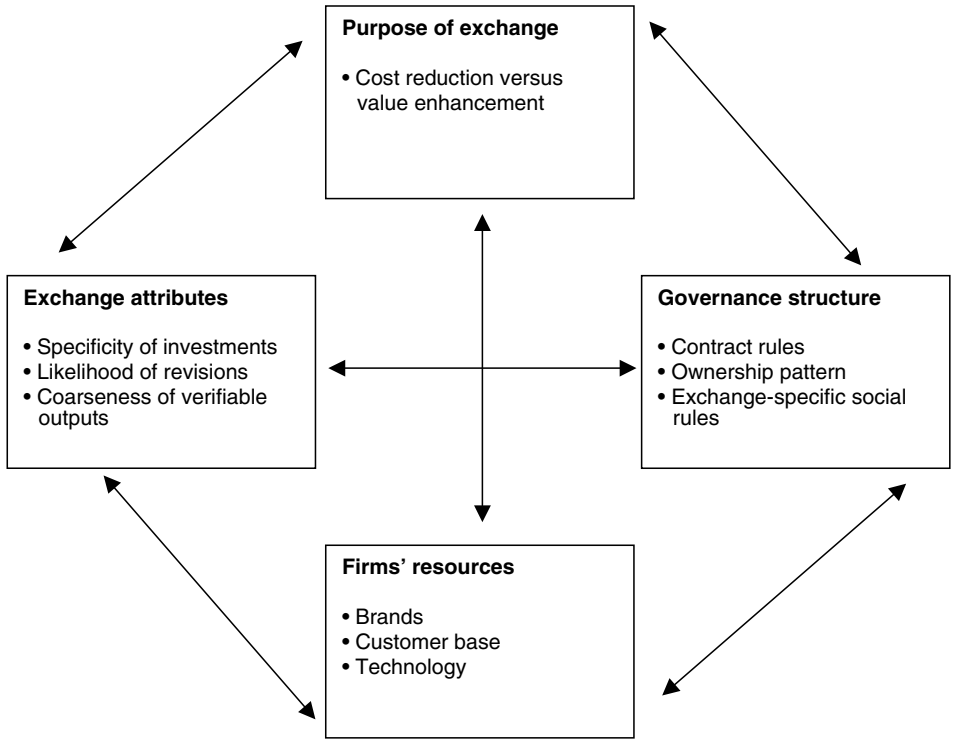
Notice that these design rules seemingly ignore strategic considerations. There are some background assumptions at work here. Most firms do not possess sufficient market power to gain from strategic ploys and feints. Thus, the cost minimization view applies to a larger set of firms. However, it is plainly incapable of dealing with instances of multiple firms organizing the same transaction in different ways. In general, one might expect that a firm with greater market power will find it more attractive to forgo efficient governance choice if this jeopardized its incumbent status.

In dealing with this matter, Williamson (1999) posed the following question as a means to fold in strategic considerations into governance design: *How should firm A, with its pre-existing strengths and weaknesses (core competencies and disabilities), organize transaction X?* Notice that two firms undertaking the same transaction may well choose different governance structures given their different capabilities.

The recent literature has responded to Williamson's challenge. Ghosh and John (1999) offered a set of design principles termed a governance value calculus. This shifts the emphasis from minimizing transaction costs to that of maximizing the value created and claimed in the exchange. This is accomplished by marrying the monopoly-efficiency trade-off to the efficiency calculus described above. This expanded approach has been distilled into a decision-making framework that we shall adapt to our pricing format problem.

The Ghosh-John model consists of the four-way match shown in Figure 7.3. The purpose of the exchange, the attributes of the exchange, the governance structure of the exchange, and the resources of the firms must all fit together in a discriminating fashion. It is the explicit consideration of the purpose of the exchange and the resources of the actors that brings strategic considerations into our calculus. Unpacking the full four-way calculus is a complicated matter,

Figure 7.3 **Governance Value Analysis**



so for most purposes, the model simplifies it by assuming that firms' resources are exogenous, at least in the medium term. It shows that the contract form employed can be explained by marrying the three governance principles articulated previously with the monopoly position (rents) of the focal firm.

Cost-Plus Versus Fixed Price Contract Forms in Procurement

We will now apply the general Ghosh-John model to the specific problem of designing price contract forms. Although price formats are but one element of the overall governance structure, the price and nonprice aspects of contracts share a particular relationship to each other. Specifically, the contract price terms are selected to support the desired product design, performance, and quantities (i.e., the nonprice contract terms). To simplify the exposition, we assume a staged process whereby the nonprice terms are assumed to be selected first, so that we can focus on fitting the price contract forms to support the previously selected nonprice terms.

Step 1: Assess Need for Price Contract Flexibility

A firm transacts with another firm to procure a product or service for monetary considerations. A product or service to be delivered is described to some degree ex ante, but a number of aspects of the product or service remain incompletely specified for several reasons.

Complexity

The complexity of a product can be summarized as the cost of specifying in sufficient (enforceable) detail its desired design, performance, and related nonprice aspects. For state-of–the-art products, and for products requiring significant research, development, and engineering, this cost can be very high. Indeed, it may be prohibitive in some circumstances. In contrast, for a commodity product, complexity is obviously much lower. Thus, the nonprice terms will be more incompletely specified for more complex products.

Clearly, if one were to lock in prices with a rigid price contract form (e.g., a firm, fixed price) in such an exchange, the costs of redesign, cost overruns in engineering and development, and so forth would all have to be borne by the supplier. It is unlikely that they would make the desired nonprice changes. Thus, the need to maintain flexibility that is expressed by the incomplete nonprice terms must be supported by equally flexible price contract formats.

Likelihood of Revisions

The likelihood of revisions to the selected nonprice terms (design, performance, quantity, timing, etc.) is much higher for all revisions in more uncertain circumstances regardless of the preciseness and description of the initial selection of these terms. Both technology and markets contribute to these uncertainties. Performance is enhanced if the initially selected terms can be revised if and when the need arises. For instance, the introduction of a technology after the initial design has been inked can be beneficial. Similarly, an unexpected competitive introduction can be countered by speeding up one's one product delivery schedule.

The need to accommodate a greater likelihood of such revisions to initial nonprice terms must be supported by the use of more flexible price contract formats. Otherwise, the seller's margins will be eroded to accommodate any costs incurred by the shift in design, quantity, and so on. Notice the asymmetry here. Revisions that decrease the seller's own costs (without a corresponding decrease in quality or other output) will be readily accommodated regardless of the price contract in force. It is the costly, but profitable, revisions that are at issue here.

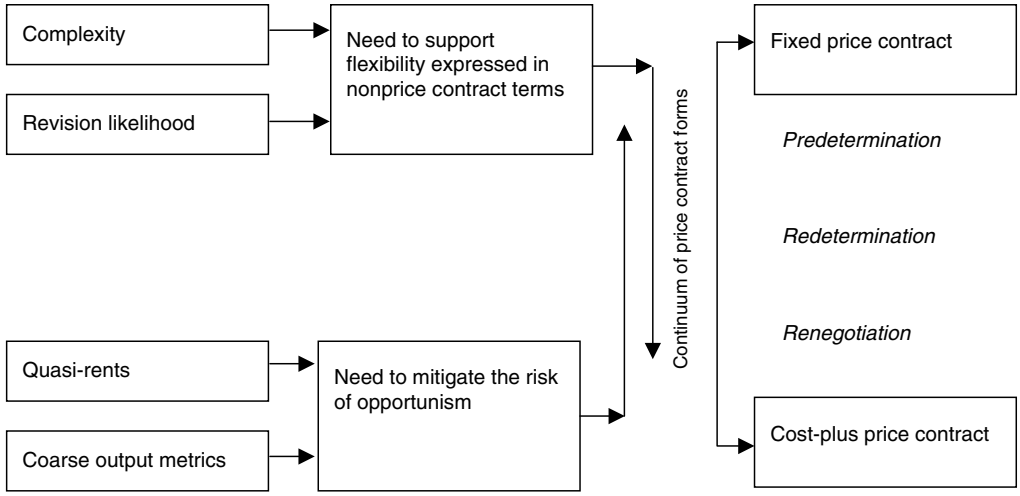
Step 2: Assess Risk of Opportunistic Renegotiation

One of the primary functions of an enforceable contract is to offer each party the security that the initial bargain will be executed. Unfortunately, not all contracts are equally enforceable, thus contracting parties have to assess the risk of opportunistic renegotiation. Opportunism here is quite different from moral hazard or shirking behavior. Most important, it refers to the practice of forcing renegotiation of incompletely expressed contract terms to increase the size of one's share of the surplus. There are two drivers of this risk.

Size of Quasi-Rents

Often, the parties to an exchange make partner-specific investments to improve quality, reduce costs, and so forth beyond the level that is realizable with general-purpose assets. These partner-specific investments generate lower returns in alternative uses, which then create quasi-rents in their current use. Each party's quasi-rents are at risk of being appropriated through opportunistic renegotiation, so the hazard can be reduced by symmetric exposure. Barring symmetric investments, the support provided by contract price terms to protecting these quasi-rents is a

Figure 7.4 **Choosing a Procurement Contract Price Format**



straightforward matter. Complete price contract forms such as firm, fixed prices are much more difficult to renegotiate to the detriment of the invested party, while cost-plus price formats can be abused in a variety of ways. Thus, the exposed quasi-rents must be supported by more rigid price contract forms.

Coarseness of Output Metrics

This describes the degree to which the nonprice terms of the contract express the desired outputs in a fine-grained fashion. For instance, the metrics available to assess the output of a contractor engaged to build one's brand is much coarser than that of another contractor engaged to sell one's product. The reduced enforceability of fine-grained measures leads one to use such coarser measures. Coarser nonprice contract terms are more vulnerable to opportunistic renegotiation. As above, this increased risk of opportunism is checked by employing more rigid price contract forms.

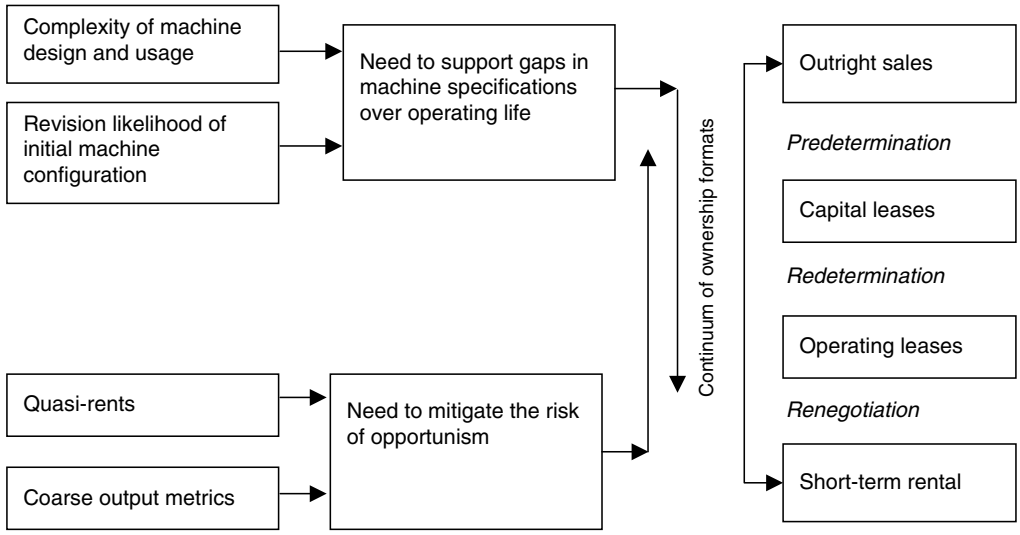
Step 3: Balancing Flexibility Against Opportunism

Figure 7.4 summarizes the drivers. Complexity and likelihood of revisions increase the need for flexibility. Quasi-rents and coarse output metrics increase the risk of opportunism. The selected price contract form is located on the continuum at a point where these two forces are balanced.

Step 4: Assess Constraints

There are two constraints on the price contract form that must be verified. First, as Appendix 7.1 shows, the price contract terms can become quite complex. This is particularly the case when the price contract form is further away from either end of the continuum. Complex indexing and renegotiation subject to constraints is not readily enforceable in all legal regimes, so the support for contract enforcement in the judiciary and polity of the host nation is a serious constraint in

Figure 7.5 **Choosing Sales Versus Leasing Formats**



this choice. Second, the constraints imposed by the cultural context must be assessed. It may be that some contract forms are not acceptable for the product or industry. This is particularly true of culturally significant products and quasi-public markets. For instance, the privatization of water utilities in several countries has been reversed because of the perception of household and agricultural water use as an essential item that should not be priced in the marketplace.

Choosing Lease Versus Sales Formats

Recall from our earlier discussion of that ownership rights altered the efficiency of contract forms. We apply this logic to the instance of leasing versus selling durable equipment. Figure 7.5 presents our template for choosing a sales price format versus a lease price format. Its logic is similar to that of the previous template for cost-plus versus fixed price contracts, except that the outcome choices are alternative ownership formats.

In contrast to the flexibility dimension used to organize alternative price contract formats in procurement, the ownership formats are organized along a rights transfer continuum ranging from the full transfer of rights at the top (outright sales) to a very narrow transfer of usage rights (short-term rentals). These alternative ownership forms are linked to systematic changes in the price determination process as well.

At the outright sales end of the continuum of ownership transfer, the net present value of the payment for the object is quite completely *predetermined* at the outset even if the payment schedule itself can be spread out in time given some financing arrangement. Such a series of payments would accompany a capital lease contract from, but its present value is nevertheless predetermined by the interest rate, timing of payments, escrows, and so forth.

However, for operating leases, the ownership is retained by the seller, so all residual asset value concerns remain with the seller. Hence, the net cash to the seller would be more incompletely known; indeed it is *redetermined* at the terminal period given the residual value of the asset. Finally, for short-term rentals, per period rental payments are likely to vary over time as the result of *renegoti-*

ated terms of multiple contracts over the life of the equipment. Thus, the tripartite classification of price determination processes identified by Crocker and Masten (1991) apply here as well.

Step 1: Assess Need for Ownership Retention

The seller is the initial owner of the durable equipment. The interest in devising an ownership format to govern the transaction can be traced to the complexity of the asset as well as the complexity of the asset in its intended use. These complexities introduce incompleteness into contract terms.

Complexity of Product and Usage

As before, the complexity of a product can be summarized as the cost of specifying in enforceable detail its design, performance, and related nonprice aspects. For state-of-the-art products, and for products requiring significant customer customization, this cost can be very high. Indeed, it may be impossible in some circumstances. In contrast, for a simpler piece of equipment, complexity is obviously much lower. The nonprice design and performance aspects will be more incompletely specified for more complex durable products as the extended operating life of durable equipment exacerbates extant gaps in specifications.

An additional source of complexity is the intended use of the equipment. In complex usage situations, it becomes difficult to pinpoint the source of breakdowns or shortfalls. It might be due to the input materials used, problems with other equipment used in tandem, or incorrect operation of the focal machine. All of these require that buyers and sellers work matters out cooperatively over an extended horizon on these nonprice aspects of the exchange.

To fix the importance of ownership formats, consider the default alternative wherein a complex, durable machine is sold outright and all ownership rights are transferred to the buyer. Post-sale requests for assistance are likely to be resisted by a seller who has no ongoing financial interest. Of course, warranties and market reputations serve to check such behavior, but as complexity grows, these governance devices are limited in their ability to overcome the problem. Suppose, however, that the seller were to retain ownership rights under an operating lease. Leasing creates an ongoing financial interest that promotes cooperative behavior by the lessor. Such cooperation is more valuable to the lessee with increased complexity. Thus, as this need grows, the correct price contract form changes to an operating lease. Under extremely high complexity, we note that even operating leases of long duration nevertheless begin to resemble a series of short-term rentals because the per-period cash flows are subject to substantial renegotiation over time.

Likelihood of Revision

The initial configuration of a durable machine is likely to be revised several times over its operating life in response to changing technology and usage. Unlike complexity, this pace of change is dictated by changes in the underlying technology as well as market needs. Machine performance (or asset value, more generally) is likely to be higher if these revisions can be made as the need arises. The need to accommodate such revisions is supported by ownership formats where the seller retains ownership of the machine (e.g., a rental or operating lease arrangement). Fundamentally, the owner of a durable asset has an improved incentive to revise a machine to maximize returns over the life of the asset. This overcomes the disabilities of transferring ownership to the buyer where the seller would be unwilling to undertake costly revisions in the absence of enforceable maintenance contracts or similar devices.

Step 2: Assess Risk of Opportunism

With durable equipment, reinvestment in asset maintenance is a costly activity that a nonowner is likely to undertake at less-than-optimal levels, which makes leases potentially hazardous. The size of this risk varies with the size of quasi-rents and the coarseness of output metrics.

Size of Quasi-Rents

The most significant problem that limits operating leases and short-term rentals is that the lessee or renter has insufficient incentives to maintain the asset. These are quasi-rents that are being appropriated by an opportunistic buyer. Costly monitoring and auditing of maintenance records might check this problem for small quasi-rents, but as the problem grows, the granting of ownership rights with a capital lease or an outright sale grow more attractive. In these latter cases, the residual value of the asset accrues to the seller, so he has the proper incentive to maintain the equipment.

Coarseness of Output Metrics

Recall that this describes the degree to which the nonprice terms of the contract express the desired outputs in a fine-grained fashion. As applied to durable equipment, this turns largely on the availability of productivity and up-time measures for the equipment. In the case of complex equipment used in tandem with other equipment, the available metrics may not be able to isolate the productivity of the focal machine. As such, the promised productivity of the equipment is likely to be expressed in much less enforceable terms, and will require good-faith adjustments by both parties. These coarser nonprice contract terms are more vulnerable to opportunistic renegotiation. As above, an increased risk of opportunistic renegotiation can be checked by shifting more of the ownership rights to the seller. Notice that ownership does not make any more information available about the machine, but only that the output metrics observed by the seller do not have to be expressed to third parties. One real-world situation that fits this scenario is equipment that will be heavily modified by the buyer. In such instances, the output metrics may be quite different from those applicable to the factory configured machine, so such transactions are better governed by an outright sale or a capital lease.

Step 3: Balancing Ownership Retention Against Opportunism

Figure 7.5 summarizes the framework. Complex equipment, combined with the likelihood of revisions to its initial configuration, increase the incompleteness of the contract, and thus the need to shift from predetermined prices to redetermined or even renegotiated prices. However, this must be balanced against the risk of opportunism because operating leases and short-term rental formats that support redetermination and renegotiation are also more susceptible to opportunism. The size of quasi-rents (asset residual value) and the availability of coarser machine productivity measures determine the risk of opportunistic renegotiation. The selected format balances these two drivers.

Step 4: Assess Constraints

There are constraints on the ownership format that can be devised. Leasing contracts are significantly affected by tax considerations. Tax authorities generally grant tax breaks to asset owners over renters. Thus, capital leases may be advantaged on this account, which may then account

for observed capital leases although the governance calculus calls for operating leases. Managers should be aware of this trade-off.

Another constraint arises from the extant legal regime. Leases are very complex contractual instruments, and with long-lived equipment, there are serious constraints imposed on such devices by uncertain judicial elements of fragile institutional environments. The blunter outright sale instrument may be more realistic than are long-run contracts requiring redetermination or renegotiation of their price elements.

Conclusion

Price formats have been relatively understudied in marketing, despite theoretical advances in the design of governance devices. In this chapter, we have examined how governance design principles can be applied to two prominent price format questions.

The first question addresses fixed price contracts versus cost-plus price contracts in industrial procurement. We created a decision framework that formulates this choice as lying on a contract continuum of pricing processes. At one end, fixed price contracts feature a predetermination process, while indexed price contracts in the middle of the continuum feature a redetermination process. At the other end, we find cost-plus contracts that feature a renegotiation process. We isolate the need for flexibility versus the need to protect against opportunism as the two opposing drivers of the right choice. Complexity and the likelihood of change increase the need for more flexibility, while the need to invest in the exchange and the coarseness of output metrics increase the risk of opportunism. The right contract balances these two effects.

The second questions addressed the issue of using contract forms in selling durable equipment. The alternatives lie on a continuum of ownership transfer. At one end, we find near-total transfer of ownership (e.g., outright sales and capital lease contracts) versus contracts that retain ownership rights with the seller (e.g., operating leases and short-term rentals). Our decision framework uses the same governance logic as above. These ownership formats are seen to lie on a continuum of price determination processes. At one end, the outright sales and capital leases employ a predetermination process, whereas operating leases employ a redetermination process. At the other end we find short-term rentals, which employ renegotiation process over successive rental contracts. Here, the need for price flexibility is driven by the complexity of the equipment and the complexity of its intended use along with the likelihood of changes in machine configuration over its lifetime. The risk of opportunism grows with machines that buyers intend to reconfigure heavily as well as with the coarseness of machine productivity measures. The correct contract balances these issues.

Appendix

Redetermination may be requested at time of deregulation and at each succeeding January 1 thereafter by seller, who will select a redetermined price from one of the following:

1. Initial price of $6.169 effective December 1, 1981, escalating monthly thereafter based on Section 102 escalation factors;
2. The price in effect immediately prior to redetermination;
3. Average of the two highest prices, selected by seller, being paid for substantially comparable gas produced in South Louisiana onshore and offshore, escalated monthly by the Section 102 escalation factor, including taxes [a most-favored-nation clause]; or
4. A price equivalent to 80 percent of the price of No. 2 fuel oil defined as 100 percent of

> the average price per MMBtu for No. 2 fuel oil as published in Platt's Oilgram for "South and East Terminals, New York Harbor District." The average fuel oil price each month will be calculated from the arithmetic average of the daily arithmetic averages of the high and low quotations for each day of the month used. To convert price per gallon to price per MMBtu, each gallon will be deemed to contain 0.138691 MMBtu [a fuel-tied provision].

Redetermined prices, including tax reimbursement, shall not exceed 110 percent of the price of No. 2 fuel oil as determined above [a maximum price provision]. If buyer, at its sole discretion, determines that the total price being paid for all or a portion of the gas is not economical, buyer may elect not to pay the price and notify seller of the price it is willing to pay. If seller is unwilling to accept such price, it can cancel the contract [a market-out provision].

Notes

1. Institutions do not refer to organizations or actors, but instead refer to the rules of interaction that govern the behavior of actors in dealing with other actors.
2. Jackson (1980) offers risk shifting as the principal driver of these choices, but this does not hold up in empirical work.
3. Clearly, bargaining can be over a specific price as in a bazaar. However, the format itself has not changed since the parties are only determining the price level.
4. Complete contracting assumes that the project or object can be satisfactorily so that court enforcement of all contract terms proceeds smoothly.
5. A financial/capital lease shifts ownership away from the seller to a third party which offers the user the capital needed to complete the transaction. These leases are largely structured to exploit deprecation tax allowances and investment tax credits.
6. Governances structures and institutional arrangements are used synonymously. We switch to the latter term to signify our emphasis on conscious design.
7. Excerpt of contract terms for natural gas supply (U.S. Department of Energy 1982).

References

Bajari, Patrick, Robert McMillan, and Steve Tadelis. 2002. "Auctions Versus Negotiations in Procurement: An Empirical Analysis." Working paper, Economics Department, Stanford University.

Bajari, Patrick, and Steve Tadelis. 2001. "Incentives versus Transactions Costs: A Theory of Procurement Contracts." *Rand Journal of Economics* 32 (3) (Autumn): 387–407.

Bercovitz, Janet, Sandy Jap, and Jackson A. Nickerson. 2006. "The Antecedents and Performance Implications of Cooperative Exchange Norms." *Organization Science* 17 (6) (November–December): 724–740.

Breyer, Ralph F. 1934. *The Marketing Institution*. New York: McGraw-Hill.

Carson, Stephen J. 2000. "Managing Creativity and Innovation in High-Technology Interfirm Relationships." Ph.D. diss., University of Minnesota.

Carson, Stephen J., Timothy M. Devinney, Grahame R. Dowling, and George John. 1999. "Understanding Institutional Designs Within Marketing Value System." *Journal of Marketing* 63 (Special Issue): 115–130.

Coughlan, Anne T., Erin Anderson, Louis W. Stern, Adel I. El-Ansary. 2001. *Marketing Channels*. Englewood Cliffs, NJ: Prentice-Hall.

Crocker, Keith J., and Scott E. Masten. 1991. "Pretia ex Machina: Prices and Processes in Long-Term Contracts." *Journal of Law and Economics* 34 (1) (April): 69–99.

Crocker, Keith J., and Kenneth J. Reynolds. 1993. "The Efficiency of Incomplete Contracts: An Empirical Analysis of Air Force Engine Procurement." *Rand Journal of Economics* 24 (1) (Spring): 126–146.

Ghosh, Mrinal. 1997. "Creating and Claiming Value in Cooperative Inter-Firm Relationships: Theory and Evidence." Ph.D. diss., University of Minnesota.

———. 2006. "An Empirical Study of Operating Leases Contracts Versus Sales Contracts for Industrial Equipment." Working paper, Marketing Department, University of Michigan.

Ghosh, Mrinal, and George John. 1999. "Governance Value Analysis and Marketing Strategy." *Journal of Marketing* 63 (Special Issue): 131–145.

Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94 (4) (August): 691–719.

Jackson, Barbara Bund. 1980. "Manage Risk in Industrial Pricing." *Harvard Business Review* 58 (4) (July–August): 121–143.

Laffont, Jean-Jacques, and Jean Tirole. 1993. *A Theory of Incentives in Procurement and Regulation*. Cambridge, MA: MIT Press.

Levy, Brian, and Pablo T. Spiller. 1994. "The Institutional Foundations of Regulatory Commitment: A Comparative Analysis of Telecommunications Regulation." *Journal of Law, Economics and Organization* 10 (2) (October): 201–246.

Masten, Scott E., and Edward A. Snyder. 1993. "United States versus United Shoe Corporation: On the Merits." *Journal of Law and Economics* 36 (1) (April): 33–70.

Milgrom, Paul. 1989. "Auctions and Bidding: A Primer." *Journal of Economic Perspectives* 3 (3) (Summer): 3–22.

Nagle, Thomas, and Reed Holden. 2002. *The Strategy and Tactics of Pricing*. Englewood Cliffs, NJ: Prentice-Hall.

North, Douglass C. 1991. "Institutions." *Journal of Economic Perspectives* 5 (1) (Winter): 97–112.

Rosen, Susan. 2004. "Norms Versus Behavior in OEM–Supplier Relationships: Antecedents and Outcomes in Intended Relational Exchange." Ph.D. diss., University of Minnesota.

U.S. Department of Energy, Energy Information Administration. 1982. "Natural Gas Producer/Purchaser Contracts and Their Impacts on the Natural Gas Market." Publication No. 0330.

Walden, Eric. 2002. "Information Technology Issues for a New Economy: Three Essays on Electronic Commerce and Information Technology Outsourcing." Ph.D. diss., University of Minnesota.

Waldman, Michael. 2003. "Durable Goods Theory for Real World Markets." *Journal of Economic Perspectives* 17 (1) (Winter): 131–154.

Williamson, Oliver E. 1999. "Strategy Research: Governance and Competence Perspectives." *Strategic Management Journal* 20 (12): 1087–1108.

ABOUT THE EDITOR AND CONTRIBUTORS

Dana L. Alden is the William R. Johnson, Jr., Distinguished Professor of Marketing at the Shidler College of Business, University of Hawai´i.

Kalpesh Kaushik Desai is Assistant Professor of Marketing at the School of Management, State University of New York, Binghamton.

Valerie S. Folkes is USC Associates Chair in Business Administration at the Marshall School of Business, University of Southern California.

Wayne D. Hoyer holds the James L. Bayless/William S. Farish Fund Chair for Free Enterprise at the McCombs School of Business, University of Texas.

George John holds the Pillsbury-Gerot Chair in Marketing at the Carlson School of Management, University of Minnesota.

V. Kumar is the ING Chair Professor, and Executive Director, ING Center for Financial Services, at the School of Business, University of Connecticut.

Jordan J. Louviere is Professor of Marketing and acting Executive Director of the Centre for the Study of Choice, Faculty of Business, University of Technology, Sydney, Australia.

Anita Man Luo is a doctoral student in marketing at the School of Business, University of Connecticut.

Naresh K. Malhotra is Regents' Professor in the Department of Marketing, College of Management, Georgia Institute of Technology.

Shashi Matta is Assistant Professor of Marketing in the Marketing and Logistics Department at Ohio State University's Fisher College of Business.

Michael A. Merz is a doctoral candidate in international marketing at the Shidler College of Business, University of Hawai´i.

Robert J. Meyer is the Gayfryd Steinberg Professor of Marketing and co-director of the Center for Risk and Decision Processes at the Wharton School, University of Pennsylvania.

Rik Pieters is Professor of Marketing and Chairman of the Department of Marketing at Tilburg University, The Netherlands.

Argun Saatcioglu is a doctoral student in the Department of Organizational Behavior at the Weatherhead School of Management at Case Western Reserve University.

Jagdip Singh is Professor of Marketing at the Weatherhead School of Management at Case Western Reserve University.

Michel Wedel is the PepsiCo Professor of Consumer Science at the Robert H. Smith School of Business, University of Maryland.