Chapter 6

A REVIEW AND COMPARATIVE ANALYSIS OF LADDERING RESEARCH METHODS

Recommendations for Quality Metrics

Thomas J. Reynolds and Joan M. Phillips

Abstract

Laddering has been used extensively within marketing to uncover the drivers of consumer decision making. Obstacles confronting laddering researchers include the time and costs of this qualitative technique as well as the lack of standard statistical measures to assess data and solution quality. In this chapter we assess the laddering research practices of both professional and academic researchers, propose a set of quality metrics, and demonstrate the use of these measures to empirically compare the traditional face-to-face interviewing method to an online one-on-one interviewing approach.

Introduction

Laddering (Reynolds and Gutman 1988) is a qualitative research technique that is widely used by marketing research professionals to successfully design and develop products (Griffin and Hauser 1982), formulate marketing and advertising strategies (Gengler and Reynolds 1995; Parry 2002; Reynolds and Olson 2001; Wansink 2003), segment markets (Reynolds 2006; ter Hofstede, Steenkamp, and Wedel 1999), and understand consumer choice and decision making (Bagozzi and Dabholkar 2000; Zeithaml 1988). The theoretical underpinnings of this technique emanate from means-end theory, which is premised on the belief that decisions are based on one's desired outcome or end-state (Gutman 1982). Central to means-end theory is the notion that individual behavior is driven by personal values, defined desired "end-states of existence" that individuals strive for in their lives (Gutman 1982; cf. Rokeach 1973).

The primary use of laddering and means-end research is to go beyond traditional techniques to obtain qualitative associations with the goal of uncovering the more personal meanings that serve as the reason that attributes and benefits (i.e., positive consequences) derive their personal salience. The networks of meaning (i.e., ladders and/or means-end chains [MECs])—obtained by probing a distinguishing attribute with some version of the "why is that important to you" question—are thought to operate top-down. That is, the higher levels are the reasons the lower levels derive their importance. For example, consider a ladder obtained from a respondent as to why a logistical transport service is viewed as preferable:

Personal value:	Success in my job (accomplishment)
Psychosocial consequence:	Feel in control of important materiel
Functional consequence:	Able to know status of materiel at all times
Attribute:	Real-time tracking system

The MECs obtained from laddering—a bottom-up interviewing technique—are interpreted as the reason the attribute "real-time tracking system" is important. In this case, the attribute, real-time tracking system, provides the worker with a long-term sense of accomplishment, which is interpreted as the respondent's motivating driver behind his/her decision to select that particular logistical service.

The ability to obtain MECs for a specific sample (e.g., decision makers for logistical services) provides a basis to develop both a lexicon of meanings (i.e., summary codes) and the explanatory connections as to why lower-level codes are important to the decision maker. The combination of the meanings *and* the connections, when summarized, provides the researcher with additional insight as to the perceptual "motivators" that underlie the customer decision-making process. As such, the laddering methodology permits the summary of this data across multiple respondents yielding a hierarchical value map (HVM), which is thought to graphically represent decision structures.

How laddering differs from other qualitative techniques is subject to debate. Proponents argue that laddering is theory driven (i.e., means-end theory; Gutman 1982) and that means-end theory is consistent with other theoretical viewpoints including both "hot cognition" (Abelson 1963) and a motivation perspective of consumer choice (Cohen and Warlop 2001). Unlike many qualitative techniques, laddering has specific rules for eliciting cognitive distinctions (Reynolds, Dethloff, and Westberg 2001) and for the coding and presentation of data (i.e., HVMs; Gengler, Klenosky, and Mulvey 1995; Gengler and Reynolds 1995; Reynolds and Gutman 1988). It is this structure that permits the direct translation of laddering research findings into positioning strategies (MECCAs model; Reynolds and Craddock 1988), and that allows for decision equity analysis (cf. Blattberg, Getz, and Thomas 2001) between target segments (e.g., heavy vs. light users) that can also be used as a basis for strategy development (Reynolds, Whitlark, and Wirthlin 2001). Critics of laddering, however, note that the manner by which value hierarchies function in consumer choice has not been empirically verified, and that some cognitive psychology perspectives seem to challenge the validity measures obtained in laddering interviews (Bagozzi and Dholakia 1999; cf. Grunert and Grunert 1995). Specifically, Bagozzi and Dholakia (1999, p. 22) contend that, "the linkages among values in the hierarchy are especially open to question if we accept the argument that (1) mental processes (e.g., the presumed personal inferences underlying means-ends connections) are not open to self-explication (e.g., Nisbett and Wilson 1977) or (2) self-knowledge is incomplete (e.g., Quattrone 1985) but instead that the linkages constitute subjective, post-hoc interpretations of previously generated responses (e.g., Dennett 1987)."

Despite its real or potential theoretical limitations, laddering has been used widely across a host of substantive research domains. Table 6.1 summarizes the breadth of laddering research applications among academics. Within marketing, perhaps the most common application of MEC analysis uses laddering to uncover the deep drivers (i.e., values) of customer decision making from which managers develop "customer-oriented" strategies (see Parry 2002 and Reynolds and Olson 2001). It is this unique ability to identify and define the drivers of consumer choice that has made laddering such an important and widely used consumer research technique. Although the procedures of the laddering methodology are outlined in detail (Reynolds and Gutman 1988; see also Reynolds and Olson 2001), there is a dearth of empirical work that has specifically ad-

Table 6.1

Empirical Laddering Studies by Substantive Area

American Presidential Politics
Bagozzi and Dabholkar (2000)
Clothing and Fashion
Botschen and Hemetsberger (1998)
Botschen, Thelen, and Pieters (1999)
Cross-Cultural
Baker, Thompson, and Engelken (2004)
Botschen and Hemetsberger (1998)
Bredahl (1999)
Dibley and Baker (2001)
Nielsen, Bech-Larsen, and Grunert (1998)
Overby, Gardial, and Woodruff (2004)
Roininenet al. (2004)
White and Kokotsaki (2004)
e-Commerce/e-Banking
Laukkanen (2006)
Laukanen and Lauronen (2005)
Subramony (2002)
Environmental Behavior
Bagozzi and Dabholkar (1994)
Bech-Larsen (1996)
Family, Children, and Parenting Behavior
Dibley and Baker (2001)
Gengler, Mulvey, and Oglethorpe (1999)
Makatouni (2002)
Manyiwa and Crawford (2002)
Russell et al. (2004a)
Søndergaard (2005)
Financial Behavior
Jain and Joy (1997)
Laukkanen (2006)
Food and Drink
Baker, Thompson, and Engelken (2004)
Baker, Thompson, and Palmer-Barnes (2002)
Bredahl (1999)
Devlin, Birtwistle, and Macedo (2003)
Dibley and Baker (2001)
Flight et al. (2003)

Goal Hierarchies
Bagozzi and Dabholkar (1994)
Gutman (1997)
Pieters, Baumgartner, and Allen (1995)
Taylor et al. (2006)
Health and Safety
Miles and Frewer (2001)
Pieters, Baumgartner, and Allen (1995)
Taylor et al. (2006)
Roth (1994)
Roininen et al. (2004)
Roininen, Lähteenmäki, and Tuorila (2000)
Higher Education
Gutman and Miaoulis (2003)
Klenosky, Templin, and Troutman (2001)
Housing
Coolen and Hoekstra (2001)
Zwarts and Coolen (2006)
Managerial Behavior and Governance
Clarke and Murray (2001)
Morandin, Bergami, and Bagozzi (2006)
van Rekom, van Riel, and Wierenga (2006)
Marketing and Advertising Strategy
Gutman (1990)
Jaeger and MacFie (2001)
Mulvey et al. (1994)
Perceived Risk
Mitchell and Harris (2005)
Product Development and Evaluation
Chiu (2005)
Grunert and Valli (2001)
Langerak, Peelen, and Nijssen (1999)
Lin (2003)
Mort and Rose (2004)
Peffers, Gengler, Tuunanen (2003)
Snelders and Schoormans (2004)
Søndergaard (2005)
Wendel and Dellaert (2005)

(*continued*)

Table 6.1 *(continued)*

Fotopoulos, Krystallis, and Ness (2003)	**Retail and Services**
Grunert and Bech-Larsen (2005)	Botschen, Thelen, and Pieters (1999)
Grunert et al. (2001)	Devlin, Birtwistle, and Macedo (2003)
Grunert and Valli (2001)	Mitchell and Harris (2005)
Gutman (1990)	Orsingher and Marzocchi (2003)
Jaeger and MacFie (2001)	Pieters, Bottschen, and Thelen (1998)
Makatouni (2002)	Vannoppen, Verbeke, and Van Huylenbroeck (2002)
Manyiwa and Crawford (2002)	**Sales and Relationship Management**
Miles and Frewer (2001)	Deeter-Schmelz, Kennedy, and Goebel (2002)
Nielsen, Bech-Larsen, and Grunert (1998)	Gengler, Howard, and Zolner (1995)
Overby, Gardial, and Woodruff (2004)	Guenzi and Troilo (2006)
Roininen, Arvola, and Lahteenmaki (2006)	Hakkio and Laaksonen (1998)
Roininen, Fillion, Kilcast, and Lahteenmaki (2004)	**Sports, Travel, and Leisure**
Roininen, Lahteenmaki, and Tuorila (2000)	Deeter-Schmelz and Sojka (2004)
Russell et al. (2004a, 2004b)	Goldenberg et al. (2000)
Søndergaard (2005)	Haras, Bunting, and Witt (2006)
Urala and Lähteenmäki (2003)	Klenosky (2002)
Vannoppen, Verbeke, and Van Huylenbroeck (2002)	Klenosky, Gengler, and Mulvey (1993)
White and Kokotsaki (2004)	Klenosky, Templin, and Troutman (2001)
Zanoli and Naspetti (2002)	Mulvey et al. (1994)
	Naoi et al. (2006)

Note: $n = 68$; some references are cross-listed.

dressed MEC quality (Grunert and Grunert 1995). Both the results of our expert survey and our review of the empirical laddering literature, reported in sections to follow, demonstrate the need for uniform quality metrics if laddering research is to reach its potential. Here we posit such a framework by proposing metrics pertaining to the reliability and validity of the MECs derived from the laddering methodology and the coding of these MECs from which the subsequent HVM is constructed. In addition to the development of quality metrics, we demonstrate the utility and effectiveness of these metrics in a field experiment that uses the measures to analytically compare two laddering methods: the traditional approach, which uses face-to-face interviewing, and an approach that uses one-on-one interviewing in an online environment. Finally, we conclude with a discussion of the implications of our findings for practice.

Laddering

Traditional laddering interviews involve a trained interviewer asking a series of questions to a respondent with the goal of abstracting the higher-order meanings that drive the respondent's perceptions and decision making. To develop an MEC for a choice discriminating characteristic, the interviewer first elicits a distinction between relevant choice options (e.g., Most preferred brand [Coke] vs. Second choice [Pepsi]), with regard to stated preference (i.e., why do you prefer Coke to Pepsi?) or actual consumption/use behavior (i.e., why do you drink more Coke than Pepsi?). Then, the interviewer sequentially probes the respondent's answers with some version of the "why is that important to you?" question, using each answer as the basis for the subsequent probe. The result of moving the respondent up the "ladder of abstraction," from attribute to functional consequence to psychosocial consequence to value, is a complete MEC, which defines both the key perceptual discriminator and the motivating reason why it is personally relevant. The number of meanings (i.e., rungs on the ladder) obtained in the laddering questioning typically range from four to six. It is the goal of the interviewer to ensure that there are meanings at each of the four levels. MECs, therefore, emerge from the aggregation of individual ladders and HVMs are constructed from the compilation of MECs. A glossary of laddering terms is presented in Appendix A.

Means-end theory posits that attribute importance is derived from satisfying an important consequence, which, in turn, ultimately derives its importance from satisfying a higher-order value orientation. As noted earlier, means-end theory may be viewed as a top-down approach to understanding choice, while "laddering," used to uncover MECs, may be seen as a bottom-up research technique (starting with attribute distinctions) whose goal is to identify the end-state that defines the motivating dynamic of the decision structure. Thus, the goal of laddering is to uncover this network of meanings, which also defines the association network of connections. Put simply, this hierarchy of meanings that comprises an MEC from a laddering interview represents the personal reasons *why* the choice discriminating attribute is important with respect to the consumer's decision making. An example of a successful laddering study involved U.S. president Ronald Reagan's 1984 reelection campaign (Bahner and Fiedler 1985; Norton 1987; Wirthlin 2004, pp. 142–45). After the Democratic convention in July 1984, Walter Mondale, the Democratic candidate, had pulled ahead of Reagan, the Republican incumbent, in some pre-election polls. In response, Reagan's reelection committee commissioned a laddering study just prior to the upcoming Republican convention to undercover the overarching values that would drive this election. The resulting HVMs allowed the research team to assess which candidate "owned" which attributes, consequences, and values, and to understand the linkages (i.e., implications) between them. To illustrate, one of the important values that emerged from the 100 laddering interviews was "Security for Oneself/ Children's Future." The HVMs showed that Mondale was most associated with (i.e., "owned") the attributes and consequences that led to this value. For example, one MEC that was "owned" by Mondale is comprised of five ladder elements consisting of the attribute "reduce poverty," which then links up to the functional consequence "fairness," which links to another functional consequence "sympathetic," which links up to the psychosocial consequence "cares and concerned about people," which links up to the value "security for oneself/children's future." Using the resulting HVMs, Reagan's research team was able to develop a now-classic television ad ("The Bear") that neutralized Mondale's hold on the "Security for Oneself/Children's Future" value, and by November was able to claim "ownership" of this important driver.[1] Although Reagan's eventual landslide victory in November 1984 may seem obvious today, in July 1984 the outcome was considerably less certain. Twenty years later, Reagan's chief political strategist credited the decisive victory to the development of campaign messages that effectively linked Reagan's leadership strengths (i.e., attributes) to the overarching values that drove the election (Wirthlin 2004, pp. 144–45). It was the laddering methodology that uncovered these key attributes and values and more importantly, the insights into linkages between them.

The Laddering Methodology

Laddering research involves a unique combination of qualitative and quantitative methods that is comprised of six distinct operations: (1) research design, (2) sample specification, (3) interviewing, (4) coding, (5) data summary, and (6) interpretation.[2] We provide a brief description of each of these operations next.

The primary component of the research design stage (1) is to determine the optimal choice-based distinctions that will be laddered to develop the MECs. Several methods can be used to elicit these key attribute distinctions (see Reynolds, Dethloff, and Westberg 2001). Triadic sorting (Kelly 1955) has the respondent describe how they think about the differences between three products (e.g., "Tell me some important ways in which two of the three brands [e.g., X, Y, Z] are the same and thereby different from the third?"). Other questioning techniques that can be used to elicit distinctions that ground decision-making understanding are: preference difference questions (e.g., "You said that your most preferred brand was X and that your second most was Y. What is it, specifically, that makes X more desirable?"), usage difference questions (e.g., "Why do you use more of brand X as opposed to brand Y?"), trend questions ("Why have you been visiting store X less/more over some time frame?"). In addition, interviewers can also use a respondent's scale judgment as an anchor to assess the discrimination between adjacent scale-points ("In terms of likelihood to purchase brand X you said 6, why not 7? Or, you said 6, why not 5?"). This latter technique to determine the most discriminating choice characteristic, termed "on the margin," provides the researcher with insight as to which one characteristic, if communicated, would have the most influence to affect choice. Finally, top-of-mind imaging asks the respondent to give one or more first-thought associations for each of several brands and then states the polarity (positive or negative) for each association. These last two elicitation techniques, top-of-mind and on-the-margin, were both used in our empirical study, described later.

As with most marketing research, laddering studies tend to use nonprobability (e.g., convenience or judgment) sampling procedures (see Bagozzi and Dabholkar 1994 and 2000 for exceptions). What is perhaps most unique to laddering is its sample specification framework (see Olson and Reynolds 2001 for a detailed explanation) that helps the researcher define the problem in decision-making terms. That is, sample specification (2) is based upon determining what groups would provide the richest understanding to address the problem of interest (e.g., loyal consumers of my brand versus those of the largest competitors, or heavy versus light users of my brand).[3] The laddering researcher specifies the sample by answering four framing questions: (i) Who are the *relevant consumers* or customers whose decisions I need to understand? (ii) For those consumers, what *particular behaviors or actions* (e.g., shopping, brand choice, consumption, or voting decisions) are most relevant to the defined problem? (iii) What are the *social and physical contexts* in which those behaviors or actions occur? (iv) What *choice alternatives* does the consumer consider when making the key decisions in those situations? Of course, balancing the sample demographically, where appropriate, to avoid bias is always a relevant consideration (e.g., Overby, Gardial, and Woodruff 2004). As with most convenience and/or judgment sampling procedures, the sample is typically pre-recruited on the basis of some combination of the aforementioned criteria for interviews to be conducted in a central location. Issues regarding geographic differences are generally overlooked to a great degree due to the prohibitive costs (financial and time) associated with interviewer travel. Interviews are rarely conducted in any more than three locations.

Interviewing (3) requires a trained interviewer to follow a general questioning protocol with the primary goal of abstracting a series of complete ladders. Once a distinction has been elicited, the interviewer ladders the respondent by taking the attribute elicited and then moving the respondent to a higher level of abstraction by asking a version of the "Why is that important to you?" question, which could include "How does that make you feel?" or "What does that do for you?" The response to that question is used as the basis for another laddering probe, again using a version of the "Why is that important to you?" question. This continues until a complete ladder is obtained, which is defined by having a verbatim response recorded at all four levels of abstraction (attribute through personal value). Interviewer expertise is required in two areas: identifying the level at which a given answer corresponds (ensuring the resulting ladder represents all four levels) and having effective questioning skills to move the respondent up the ladder of abstraction.

Note that it is not uncommon for respondents to provide two sequential comments at the same level of abstraction (e.g., both functional consequences), or to, say, follow a psychosocial consequence with a lower-level reply (e.g., functional consequence). When this happens, laddering protocol calls for the interviewer to use probes to elicit the higher- (or lower-) order responses needed to complete the ladder. Therefore, it is possible for the "length" of a complete ladder to consist of more than four elements (but not fewer). Although less common, some respondents begin at a higher-order level. When this happens, interviewers should first use probes to "chute" the respondent down to a lower level of abstraction (e.g., "What is it about brand X that makes you feel secure?") before using probes to "ladder" back up to higher levels of abstraction (see Reynolds, Dethloff, and Westberg 2001; Reynolds and Gutman 1988). These scenarios also illustrate the unique skill required of laddering interviewers to accurately identify the levels of abstraction of the respondents' verbatim comments within the MEC, and to use the proper probes to ensure that a complete ladder is elicited.

Coding (4) involves the content analysis of the verbatim responses obtained in the laddering interviews, based on common meanings at the respective levels of abstraction (Kassarjian 1977). The first step in this operation is to inspect the ladders for completeness and to develop a set of summary codes that reflect the elements mentioned. This is done by classifying each element (i.e., response) into the four basic levels—attribute, functional consequence, psychosocial consequence, value—and then assigning a summary code to each element. The objective is to focus on selecting codes with meaning central to the purpose of the study as it is the relationships between these elements that are of interest and not the elements themselves. Once the master codes have been finalized and numbered, these numbers are used by two or more coders to score each element in each ladder, and measures of intercoder agreement and/or reliability (e.g., Perreault and Leigh 1989) should be assessed to ensure coding quality. The end result of this operation is a "score matrix" with each row representing an individual ladder with the sequential element within the ladder corresponding to the consecutive column designations (see Appendix B for a hypothetical score matrix and Reynolds and Gutman 1988 for a complete description of this procedure). This lexicon of meanings is one of the key outputs of laddering, which, for the marketing manager, serves to translate the manufacturer's orientation into consumer language. Importantly, this qualitative richness also provides a valuable input for strategy specification and the subsequent creative development process.

Summarizing the data (5) involves two types of analyses. The first is the construction of a summary decision map (e.g., hierarchical value map: HVM), which is accomplished by computing the number of direct (A directly precedes B) and indirect (A indirectly precedes B) implications across all of the codes. A "significant" threshold value is selected, which defines meaningful implications between the four "levels of abstraction" (i.e., attribute → functional consequence → psychosocial consequence → value) to be represented in the decision map. The determination of the threshold value typically captures 70 percent of the implicative connections in the data (Gengler and Reynolds 1995; Reynolds and Gutman 1988), avoiding the smaller, infrequent connections that are viewed as idiosyncratic (see Appendix B for an example of an implication matrix and construction of an HVM). The decision map, then, is interpreted as a representation of the primary decision structures in the research domain (e.g., brands, services, candidates, etc.) of interest (Gutman 1982; Walker and Olson 1991), with each pathway (i.e., linkages from bottom [attributes] to the top [values]) representing a possible decision network. The second type of analysis involves contrasting various predetermined sample breaks in an attempt to understand the basis for the "decision differences" (e.g., brand A vs. brand B "loyal customers," or heavy vs. light users), which serves as additional input relative to the competitive environment. This is usually based upon the frequency of occurrence of the codes by the respective sample groups, using the overall decision map as the basis to summarize the differences (Reynolds, Dethloff, and Westberg 2001). The common goal of all of these analytic procedures is to provide an interpretive framework (6) from which to develop alternative strategic positioning strategies (e.g., Parry 2002; Reynolds and Olson 2001).

The steps in the laddering process require the researcher to make many judgments, from how the problem is initially framed, to the resulting sample, to what analysis criteria to base the solution upon. This hybrid approach to taking what is inherently qualitative data and applying additional methodological rigor to yield semi-quantitative decision representations is clearly fraught with potential sources of error. Thus, classifying laddering primarily as a qualitative methodology is appropriate. However, two specific areas of the research process that can be quantitatively assessed are the quality of the initial laddering interview data and the quality of the coding process. Given that error from either source would compound any of the subsequent analytical laddering procedures, the value of obtaining metrics to assess ladder quality would seem to be fundamental to laddering research. To assess the practical importance and potential of these quality metrics, we provide an overview of the practice of laddering research by reporting on the results of an expert survey of laddering research professionals and an integrative review of the academic marketing literature.

The Practice of Laddering Research

Laddering does present the researcher with certain obstacles; chief among them is the laddering interview. As with any qualitative method, obtaining valid measures of higher-level constructs can be problematic and the degrees of freedom inherent in laddering pose additional challenges, particularly for the interviewers. Laddering has traditionally required highly trained interviewers to follow a system of established, albeit fluid, procedures to ensure the elicitation of complete ladders that consist of valid responses at each of the four levels of abstraction.

It is this fluidity of the interviewer protocol that uniquely sets laddering apart from traditional attitudinal and satisfaction surveys (e.g., Schuman and Presser 1996). Although both laddering and attitudinal research strive to capture the latent drivers of behavior, attitudinal surveys have traditionally been standardized with predetermined response alternatives so as to minimize interviewer and coding error. Whereas standardized responses may indeed minimize both types of error, this approach also limits the cognitive effort and subsequent involvement of the respondent in searching for the most appropriate and personally meaningful answer. The problem with this simplified method is that it frequently results in more superficial responses and biased answers, particularly if the questions themselves are not well formulated (Bradburn, Sudman, and Wansink 2004, pp. 158–59). This is also a limitation of "hard" laddering approaches (e.g., Walker and Olson 1991), particularly those that use self-administered questionnaires with standardized responses (e.g., ter Hofstede et al. 1998), as these techniques make it impossible to detect strategic processing by the respondent (Grunert and Grunert 1995).

As the following example from the 2004 U.S. presidential election illustrates, this problem can be compounded in surveys asking about higher-order personal values. After the polls closed, the media reported large discrepancies in results from exit polls asking about the role of "moral values" on the voters' choice of president. Later analysis revealed that the disparate findings were related to the different response formats used in the surveys, with a substantial increase in the importance of “moral values” reported in the poll using a standardized fixed-choice format over the poll using an open-ended format (Babington and Faler 2004; Menand 2004).

As the above case in point demonstrates, although the benefits of in-depth laddering interviews over traditional surveys for yielding a superior understanding of consumer decision making are clear, the associated costs can be significant. Beyond the need for experienced professionals for the framing and design function, the costs of hiring, training, and compensating skilled laddering interviewers, coders, and analysts, as well as respondent costs, are sizable. The average cost per interview usually exceeds US$1,000, which is significant given that many laddering studies require 80 to 150 interviews. A secondary cost that can pose an even greater barrier is the time (usually 10 to 12 weeks) required to complete a laddering study. Undoubtedly, these costs have required some businesses to forego the use of laddering research to address problems to which it would have been well suited, while requiring others to take procedural shortcuts that could compromise the validity of their findings.

Expert Survey of Laddering Research Professionals

To assess the above conjecture, identify other obstacles, and get a general sense of laddering research as commercially practiced, we conducted a small expert survey of ten marketing research firms recognized as having considerable expertise in conducting means-end research. The research firms were selected by a judgment sampling procedure based upon our prior knowledge, an Internet search, and an online search of the qualitative marketing research firms in the American Marketing Association’s Marketing Services Directory (2004). Search terms included “means-end research” and “laddering.” Of the ten firms contacted, nine agreed to participate. These phone interviews, involving seven questions, were audio-recorded and transcribed.

Interviews lasted approximately 25 minutes and the informants, all of whom held the title of Principal and/or Research Director, had between 2 and 21 (mean = 12.8) years of experience conducting laddering research. Informants received their training in means-end research and the laddering methodology in their graduate coursework, at professional workshops, and/or through reading academic literature (e.g., Reynolds and Gutman 1988). Their firms had conducted an average of 5.3 laddering research projects in the prior twelve-month period. When asked whether the number of laddering projects in the past year was steady or reflected a trend up or down from previous years, 67 percent responded that the number of laddering projects had been steady, while 22 percent noted an upward trend and 11 percent indicated a slight downward trend. Informants were also asked to name some of the more interesting research problems that laddering had been instrumental in solving. Here, informants described with great enthusiasm their own successful projects that addressed such problems as developing new products, branding, positioning, developing communications strategies, and understanding consumer choice and loyalty. These projects were conducted within a diverse group of industries that included pharmaceuticals, consumer packaged goods, industrial products, financial services, cigarettes, gaming, plastics, and politics.

Eight of the nine informants (88.9 percent) considered project cost to be a major obstacle to conducting laddering research. Although all agreed that the average cost per interview (with respect to total project cost) varied widely, most informants estimated the per-interview cost to be between US$1,000 and $2,000. One informant noted that data collection (field costs and interviewing) alone can account for as much as 40 percent of the entire project cost. Other obstacles to laddering research mentioned included: the time required to complete a project, the ability to hire quality interviewers and analysts, the demands associated with coding, and the management of client-relationship issues regarding the necessary rigor of the research process to obtain the desired strategic insight. Interestingly, there was a presumption (both stated and implied) by nearly all informants that overcoming these obstacles necessitated “shortcuts” or “flexibility” in the application of laddering procedures. In fact, some informants readily assumed that maleficent shortcuts were taken by competing firms. However, this was perhaps most evident when informants were asked to describe their own in-house operational procedures for three laddering research tasks: questionnaire design, interviewer training, and coding. In addition to describing their operational procedures in detail, most informants voluntarily acknowledged that their procedures often fell short of “proper” or “textbook” laddering.

The severity of these procedural breaches ranged from minor to serious. Three of the more serious violations mentioned included: (i) using overly structured questionnaires that did not allow for probing in order to compensate for inexperienced interviewers or to allow for noninteractive computerized interviewing, (ii) interviewing groups of people instead of individuals, and (iii) not content analyzing and coding the verbatim responses but constructing ladders and HVMs based solely on the researcher’s judgment after listening to the audio recordings of the interviews.

Nearly all informants echoed a similar sentiment; namely, that strict adherence to classic laddering procedures required a rigor that was difficult, if not impossible, to achieve in practice.[4] Although this finding was not particularly surprising, we were struck by the degree of contrition expressed by some informants when making their unsolicited confessions. It was evident that these expert researchers were very much aware that compromising procedural rigor can hinder data quality, thereby hampering their ability to construct valid and meaningful HVMs. However, there was a consistent overarching belief that the benefits of laddering research appreciably outweighed the costs associated with these procedural infractions. It seems that the tacit assumption underlying this belief was that the laddering technique is robust. Perhaps the foremost indicator of this assumption is that none of the informants mentioned conducting reliability checks to their coding, or assessing the quality (i.e., completeness) of their ladders.

Summary of Expert Survey Findings

In sum, our findings suggest that: (i) the expert researchers surveyed are highly knowledgeable about laddering and well-trained in its use; (ii) they especially value means-end research and the insights into consumer decision making that the laddering technique reveals for their clients; (iii) the greatest obstacles in conducting laddering research are the time and costs of interviewing and coding, along with finding qualified interviewers; (iv) these obstacles frequently necessitate methodological shortcuts that may hinder ladder quality and interpretation; (v) these shortcuts are permissible, based on a tacit assumption that the laddering methodology is robust and provides the insights they require, despite the shortcuts; and (vi) there was little to no use of validity checks to verify MEC quality. Finally, we suggest that implicit in the informants’ confessions was the desire to service their clients’ research needs without the cost and time burdens associated with the laddering interview, while also ensuring the quality of their ladders and ultimately their findings and strategic recommendations.

A Review of Empirical Academic Laddering/Means-End Research

Laddering also is used by many academics for substantive application to problem solving. A December 2006 online search of the Social Science Citation Index (SSCI) listed 137 citations of Reynolds and Gutman (1988), considered by many academics and practitioners to be the classic "how to" article on the laddering research technique. Additional searches on Google Scholar yielded another seventy citations in academic journals. Also included are five articles appearing in an October 1995 special issue of *International Journal of Research in Marketing* on means-end research and not found in the online searches. ter Hofstede et al.'s (1998) Association Pattern Technique (APT), discussed below, is another means-end research method. A further online search for articles citing this work yielded two additional articles that did not also cite Reynolds and Gutman (1988). Although there are limitations associated with online literature searches (Cooper 1998), and undoubtedly the product of our search understates the actual number of articles published, we do believe that this collection of 214 articles represents a reasonable sample from which to draw inferences about how and why laddering research techniques are used by academic researchers. Of these 214 articles, we were able to obtain copies (in English) of 205 (95.8 percent). We content analyzed each of these 205 articles, and based on this analysis developed a four-group classification: conceptual references, technical references, methodological developments, and empirical laddering/means-end research. An article was classified into one of these groups based upon the purpose of the laddering reference (e.g., Reynolds and Gutman 1988) within that work. All 205 articles were successfully classified into one of the four groups. To gain a better understanding of the role of laddering within academic marketing research we briefly review the articles within each of these four groups.

Conceptual Reference Group (18 Percent)

Although Reynolds and Gutman (1988) is a methodological article that describes a research technique, 18 percent (*n* = 37) of the articles that made reference to this work cited it to support or develop a conceptual framework. The articles in the conceptual reference group typically used the laddering citation as support for a conceptualization that posits that underlying a consumer's choice or behavior is a cognitive structure or "network of reasonings" (e.g., Sirsi, Ward, and Reingen 1996). This structure is hierarchical in that behavior derives its meaning from the higher levels of abstraction (e.g., Dabholkar 1994). Positioned at the top of this cognitive hierarchy are terminal values, believed to be the ultimate motivators of consumer behavior (e.g., Kamakura and Novak 1992; Thompson and Troester 2002). Consistent with these examples, our review of the articles within the conceputal reference group found that marketing academics frequently operationalize means-end theory as laddering (cf. Guiltinan 2002).

Technical Reference Group (43 Percent)

Articles making technical reference to the laddering methodology comprised our largest group (*n* = 88, 43 percent). Here, the laddering citations typically referred to the availability of the laddering technique itself (e.g., Lageat, Czellar, and Laurent 2003), that it has been used by others in the past (e.g., Veryzer 1999), that it may be used as an alternative to the technique or techniques being used in the work that cites it (e.g., Zaltman and Coulter 1995), or that it is a technique that researchers may want to consider for a specific marketing problem (e.g., product development; Griffin and Hauser 1993; Veryzer 1998). Also included in this group are articles that demonstrate an application of the laddering technique to show how laddering can be used by practioners to, say, develop advertising or marketing strategy (e.g., Reynolds and Craddock 1988; Reynolds and Whitlark 1995), and articles that introduce the laddering method to nonmarketing disciplines (Tan and Gallupe 2006). Finally, this group also included articles that did not quite "qualify" for a classification in the empirical group below.

Methodological Developments Group (6 Percent)

Just 6 percent (n = 12) of the 205 academic articles focused on suggesting improvements for the analysis of laddering research data or the laddering methodology itself. Suggested analytical improvements include data analysis software (e.g., LADDERMAP; Gengler 1995; see Lastovicka 1995 for a review), heuristics for selecting cutoff levels (Leppard, Russell, and Cox 2004; Pieters, Baumgartner, and Allen 1995), and methods to improve the graphic representation of HVMs (Gengler, Klenosky, and Mulvey 1995; Gengler and Reynolds 1995). Others have proposed various supplemental analyses as ways to enhance the interpretability of laddering research results. These supplemental analyses include cluster analysis (Aurifeille and Valette-Florence 1995; Lin and Yeh 2000), correspondence analysis (Valette-Florence and Rapacchi 1991), network analysis (Pieters, Baumgartner, and Allen 1995), and nonlinear canonical analysis (Valette-Florence 1998).

Those suggesting improvements to the methodology itself have proposed what may be deemed an efficiency paradigm (e.g., Grunert and Grunert 1995; ter Hofstede et al. 1998; cf. Walker and Olson 1991) given that the suggested improvements would result in the "hardening" of traditional laddering techniques. Indeed, several of the empirical articles (discussed next) referred to "hard" and "soft" laddering approaches, a designation first introduced by Grunert and Grunert (1995). "Soft" laddering approaches restrict the respondent's natural flow of speech as little as possible, whereas "hard" laddering refers to techniques "where the respondent is forced to produce ladders one by one, and to give answers in such a way that the sequence of the answers reflects increasing levels of abstraction" (Grunert and Grunert 1995, p. 216). Grunert and Grunert conjecture that soft approaches are potentially better when respondents have overly weak or elaborate cognitive structures, as in cases of high or low involvement and experience. In contrast, harder approaches may be best for cases of average involvement and experience, but these "hard" approaches make it impossible to detect strategic processing by respondents.

The "hard" approach used most often is a paper-and-pencil approach that employs a questionnaire format. Usually, respondents write their answers to the "why is it important to you?" questions in boxes, with each level of abstraction having its own box (e.g., Walker and Olson 1991). Perhaps the "hardest" of laddering approaches is ter Hofstede et al.'s (1998) Association Pattern Technique. This paper-and-pencil approach uses a standardized fixed-choice format amenable to a mail survey, in which the respondents check boxes of predetermined attributes, consequences, and values from which a three-level ladder is constructed. One of the drawbacks of this technique is that there is no distinction between the functional and psychosocial consequences—the latter of which is most relevant for developing advertising strategies (Reynolds and Trividi 1989). Although this method offers considerable cost savings over traditional laddering, and has been used for international market segmentation (ter Hofstede, Steenkamp, and Wedel 1999), thereby overcoming geographic constraints usually associated with traditional laddering, it is subject to the same limitations of using standardized responses in attitudinal surveys discussed earlier.

There has been much discussion (e.g., Grunert, Beckmann, and Sørensen 2001; Grunert and Grunert 1995; Phillips and Reynolds 2009) and some empirical work directed at assessing how the mode of administration (e.g., face-to-face vs. paper-and-pencil vs. "hard" computerized presentations) might impact the laddering research results (Russell et al. 2004a; Russell et al. 2004b). Although differences between methods were found using a common cutoff and subsequent inspection of HVMs (i.e., the soft laddering map was more complex), without uniform quality metrics to assess data and coding, it is difficult to make meaningful comparisons between laddering approaches (cf. Huber, Beckmann, and Herrmann 2004).

One reason, perhaps, for the lack of work addressing methodological issues related to laddering is the subjective nature of many of the decisions that need to be made during the research process. The result is a methodology with a strong bias to the expertise of the researcher and few standards of comparison to evaluate the quality of the research. When considered in light of other widely used research methodologies, it is possible to conclude that laddering suffers from a lack of standard statistical metrics needed to evaluate the quality of its key ingredient: namely, the interviewing and subsequent coding of the ladders (cf., Grunert, Beckmann, and Sørensen 2001).

Empirical Laddering/Means-End Research Group (33 Percent)

One-third (n = 68) of the articles were classified into the empirical laddering research group. The criterion used for this designation was that the article had to report a complete laddering study that included a description of research design, procedures, and results of analyses that aggregated individual ladders into MECs. Not included were studies that reported qualitative descriptions of a series of individual ladders, individual causal maps, portions of laddering studies used to illustrate an application, or qualitative depth interviews that used an adapted version of the laddering interview technique; these articles were classified into the technical reference group.

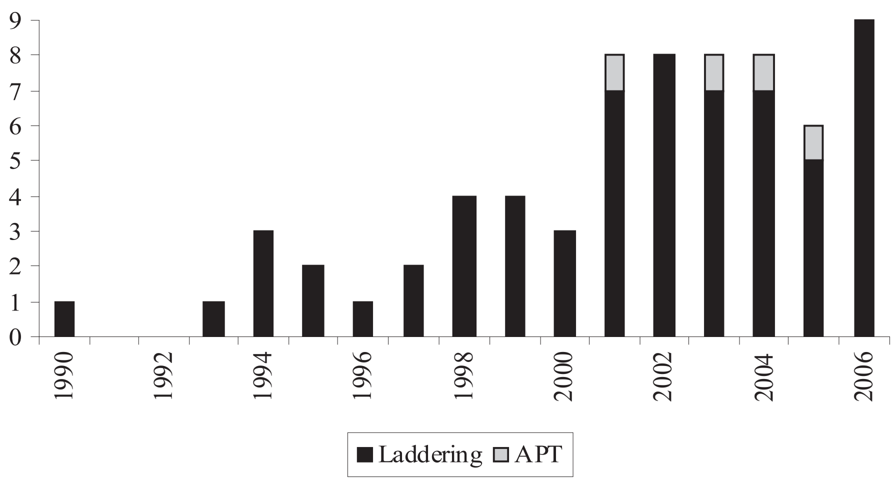
Our search found 68 empirical laddering/means-end research studies published in 40 different journals, with the first publication appearing in 1990. Figure 6.1 depicts the primary outlets for laddering and means-end research and the number of articles published by each journal between 1990 and late 2006. Figure 6.2 shows the number of laddering articles published by year. There appears to be a jump in the number of articles appearing after 2000. It is not possible to tell if this rise is an artifact resulting from online search procedures (i.e., access to electronic journal content not beginning until 2001) or an indication of increased interest in means-end research. Note also that APT, ter Hofstede et al.'s (1998) more recent "hard" laddering approach, has contributed to this rise at the rate of about one publication per year. Finally, observe that after a rise in 2001, the rate of empirical laddering/means-end research by academics has been rather steady at about eight articles per year. This finding is consistent with the perspective held by the practioners in our expert survey—namely, that laddering research is in a "steady state" with possible indications of a slight upward trend.

Table 6.1 presents these 68 empirical studies classified by substantive area. Note the rich array of topics addressed by these studies, and number of articles within each area. Given that laddering was developed primarily to study consumer behavior, perhaps it should not be surprising that the substantive domain for 38.2 percent of the articles is food and drink. This emphasis on studying the consumption and marketing of food products is also reflected in Figure 6.1, where we see that journals specializing in the topical area of food research are also the major outlets for empirical laddering studies.

Although other marketing research methods have been used to study the same substantive areas listed in Table 6.1 (see 132–133) an analysis of the stated aims of these empirical articles provides insight into why laddering was used to achieve these objectives. In general, the stated research objectives were (1) to understand the motivational aspects of consumer choice and decision making, and (2) to understand the hierarchical cognitive structure (i.e., the way the consumption-relevant knowledge is stored and organized in memory) associated with the decision (cf. Grunert and Grunert 1995). By understanding consumers' motivations and their decision-making structure as reflected in the resulting HVMs, it is possible for the researcher to (3) compare these revealed motivations and structures across different groups of consumers. For example, Gengler, Mulvey, and Oglethorpe (1999) investigated mothers' choice to breastfeed infant children and compared

Figure 6.1 **Laddering/Means-End Empirical Research Articles Published by Journal**

Figure 6.2 **Empirical Laddering/Means-End Articles Published by Year (n = 68)**



the HVMs of mothers who continued to breastfeed with those who chose to stop. The authors note that the results of this study could be used to develop public service campaigns that reinforce the benefits of breastfeeding. Similarly, in an effort to better understand cross-national differences that consumers have with respect to wine, Overby, Gardial, and Woodruff (2004) compared the HVMs of American and French consumers. Interestingly, cultural difference appeared to affect the consumer value hierarchy at the consequence level more than at the attribute level. Table 6.1 provides several additional examples of empirical laddering research that was conducted to better understand differences across consumer groups and cultures.

Given the seeming importance of culture to laddering research findings, we examined where the research for these 68 articles was conducted. The plurality of this research was conducted in Europe (n = 39, 57.4 percent), while 29.4 percent (n = 20) was conducted in North America, and 4.4 percent (n = 3) was jointly conducted within both continents. In addition, a much smaller percentage of laddering research was conducted in both Asia (n = 3, 4.4 percent) and Australia (n = 3, 4.4 percent).

Perhaps more important to its findings than where the research was conducted, is how it was conducted. Specifically, what laddering approach was used? Over two-thirds (n = 46, 67.6 percent) of the research was conducted using the traditional face-to-face approach—a "soft" laddering technique. Semi-soft techniques were also used as 5.9 percent (n = 4) of the studies used a telephone approach and 1.5 percent (n = 1) used a series of focus groups. A "hard" paper-and-pencil approach (e.g., Walker and Olson 1991) was used in 16.2 percent (n = 11) of the studies, while 5.9 percent (n = 4) of the studies used ter Hofstede et al.'s (1998) APT approach. In addition, 2.9 percent (n = 2) used a mixture of hard and soft approaches (e.g., Russell et al. 2004a).

As one would expect, sample size is not independent of laddering approach ($F(3, 65) = 10.90$, $p < .001$, $\eta^2 = .35$). Although the average total sample size is 98.1 (SD = 104.8), sample sizes are positively related to the "hardness" of the laddering approach: face-to-face = 63.2 (SD = 76.4), phone = 186 (SD = 176.9), paper-and-pencil = 148.6 (SD = 92.2), APT = 283.0 (SD = 103.9). It is worth noting that the sample sizes reported above are the total number of respondents used in the articles. As we described above, analyses usually entailed "splitting" these total sample sizes into two or more groups for purposes of comparison.

Coding quality measures such as inter-coder agreement and/or reliability were reported in just 40.6 percent of the applicable articles (coding is not usually needed with APT) despite the fact that procedural descriptions consistently made mention that the responses were coded, usually by more than one coder. Other potential quality indicators that were reported by the 64 non-APT studies were: implication matrices (26.6 percent), cutoff levels (73.4 percent), and total implications accounted for (i.e., explained variance, 40.6 percent). However, only total implications accounted for can be meaningfully interpreted in an absolute sense when this metric is reported by itself. Of those reporting this metric, the proportion total (direct and indirect) implications accounted for averaged .71 (SD = .16), which is consistent with the proportion recommended (i.e., ~.70) by Reynolds and Gutman (1988) and others.[5] A couple of articles reported just the proportion of direct implications accounted for (e.g., Overby, Gardial, and Woodruff 2004 reported that .32 of direct implications were accounted for).

In addition, a few authors made mention of average ladder length and one-third reported that the analyses used LADDERMAP software (Gengler 1995). No article assessed the completeness of its ladders, although all 64 studies (100 percent) reported one or more HVMs. In addition, some articles used additional analytical techniques to supplement the laddering analyses. These included social network analysis (20.3 percent, e.g., reporting abstraction ratios and measures of centrality and prestige; e.g., Pieters, Baumgartner, and Allen 1995; Morandin, Bergami, and Bagozzi 2006; Taylor et al. 2006), correspondence analysis (9.4; e.g., Bredahl 1999; Roth 1994), and cluster analysis (4.7 percent; e.g., Botschen, Thalen, and Pieters 1999; Klenosky, Gengler, and Mulvey 1993).

Overall, these findings indicate that there is a need for greater consistency in the manner in which laddering research is reported. Only one study (van Rekom, van Riel, and Wierenga 2006) attempted to assess the reliability of the laddering data by calculating a coefficient of agreement using split-half samples (see Kassarjian 1977). Clearly, the availability of uniform quality metrics would provide researchers with effective means to compare and integrate research findings and to compare the effectiveness of "hard" and "soft" laddering approaches.

Summary: The Problem

Although laddering procedures are well specified, they entail a rigor that researchers often find prohibitive. For competitive reasons, many professional and some academic researchers have adopted a *quasi*-laddering approach by taking procedural shortcuts on the presumption that the methodology is robust. Yet, there is little empirical evidence to support this presumption. Moreover, in contrast to other widely used methodologies, we know of no commonly accepted and/or utilized statistical metrics that permit rigorous comparisons across procedures and approaches. Furthermore, as the construction of hierarchical value maps (HVM) is dependent on coding consistency and completeness of the ladders, this assumption is problematic. This fact, in combination with the small sample sizes typical to laddering research, thus highlights the need for standard reporting statistics to facilitate the evaluation of the reliability and validity of these research findings.

Proposed Resolution

In this section, we propose three solutions intended to directly address the problems noted above. First, we focus on the issue of assessment, by specifying a set of quality metrics and methods. Our objective in proposing these quality metrics is that they become the standards by which professional and academic laddering research is conducted and evaluated. In essence, we are proposing that an additional laddering operation, quality assessment, be added to the six procedural steps outlined earlier. Our hope is that these statistics are reported for all laddering studies, to aid both researchers and their clients in evaluating the quality of the research and ensuring the validity of its findings.

Second, we evaluate a new method that involves conducting laddering interviews online using voice-over-net technology for questioning. This online approach addresses the cost and efficiency constraints previously mentioned, as well as the geographic limitation of sampling from only a few locations. This evaluation is done by describing the results of a joint laddering study that uses an identical design and sample specification, but that used two different interviewing approaches: (1) traditional face-to-face and (2) one-on-one online.

Third, to demonstrate the utility and effectiveness of the proposed quality metrics we conduct a comparative analysis of these two approaches. The combination of these evaluations should provide a meaningful perspective from which to assess the viability of conducting laddering interviews via the Internet.

The overarching issues, then, that need be addressed to provide quality standards for laddering research are: (1) develop a standard measure of coding quality and reliability, (2) develop a metric to assess the quality of the laddering data, and (3) develop reliability and validity measures for comparing alternative laddering methodologies.

Quality Metrics

Coding Quality

Evaluating the quality of the coding process requires a summary measure of coding quality (C_Q), which can be obtained by having multiple coders independently assign codes and then comparing their lists. For simplicity, we suggest C_Q equal a straightforward percentage of agreement between two coders on assigning the same code to the verbatim comment (i.e., ladder element). This first requires developing a set of common codes, by level, as a prior step to the coding process. Once the discrepancies in the coding are identified, they must be resolved, either by discussion or by a third-party judge. It is also important to note that coder agreement is just one facet of coding reliability. The number of codes, which for many laddering studies will be ample, should also be considered when assessing the reliability of coder judgments of nominal data (Perreault and Leigh 1989; Rust and Cooil 1994). We use Perreault and Leigh's reliability index (I_r) in the following analyses. Table 6.2 presents the formulas for these proposed quality metrics and those discussed below.

Ladder Quality

The goal of a laddering interview is to obtain a complete MEC that encompasses elements at all four levels of abstraction. The value in obtaining a complete ladder is its applicability. That is, positioning strategy requires knowing what associations tie ideas together and the more associations that can be labeled and included in the analysis, the more productive is the research. Therefore, if a respondent skips a level but has it in her mind already, it is worth obtaining for the sake of completeness. If it is not already in mind, it is worth finding what meaning(s) bridge the two elements in her ladder. Consequently, the quality of the interviewing process is reflected in the number of ladders that have a complete set of elements. Thus, ladder quality (L_Q) is a simple percentage statistic, which provides a key MEC-quality metric that reflects the percentage of ladders that succeed in meeting this basic criterion.

Reliability and Validity Measures for Comparing Laddering Methodologies

The construction of a summary hierarchical value map (HVM) from MEC data requires selecting an appropriate threshold cutoff, usually one that accounts for approximately 70 percent of the total (direct + indirect) implications. It is this threshold level that defines the "significant" connections to be mapped. "Significant" connections reflect those "code-pairs" meeting or exceeding the threshold value.

The dilemma, when constructing an HVM from MEC data, is that comparing two MEC solutions, by analyzing their respective HVMs, is subjective and does not directly lead to an objective quantifiable assessment. To resolve this problem, the structure of the HVM (computed from the implicative relations between codes) is considered with respect to the corresponding dominance relations, between codes (Coombs 1964). The definitions and relationships of dominance relations with respect to quantifying the key aspects of tests, reliability and validity, are developed in Ordinal Test Theory (Cliff and Keats 2003, ch. 6). This approach to partitioning variance (defined by dominance implications) permits the computation of parallel summary statistics defined by Classical Test Theory. Adopting this dominance-analysis framework provides for a conceptual translation of MEC data, from which these types of variance may be derived to produce estimates of reliability and validity. The measures developed here should most accurately be termed *quasi*-reliability and *quasi*-validity, in that the metrics do not possess any strong statistical properties. Rather, they are estimates that are conceptually derived to capture the theoretical basis of both measures.

To assess *quasi*-reliability, or r_q, each set of MEC data is summarized by the number of connections, or implications, between codes. As such, the total variance may be thought of as the sum

Table 6.2

Proposed Quality Metrics

Quality Metric	Symbol	Formula
Coding quality	C_Q	$\frac{\text{No. of ladder elements with same code}}{\text{all ladder elements coded}}$
Coding reliability index (Perreault and Leigh 1989) n_c = number of codes	I_r	$\left[\left(C_Q - \left(\frac{1}{n_c}\right)\right)\left(\frac{n_c}{n_c - 1}\right)\right]^{\frac{1}{2}}$
Ladder quality	L_Q	$\frac{\text{No. of complete ladders}}{\text{all ladders}}$
Quasi-coefficient of determination	r_q^2	$\frac{\text{Total “significant” implications*}}{\text{Total (direct + indirect) implications}}$
Quasi-reliability	r_q	$\sqrt{\frac{\text{Total “significant” implications*}}{\text{Total (direct + indirect) implications}}}$
Quasi-validity	v_q	$\frac{\text{Total “common, significant” implications* with comparison data set}}{\text{Total (direct + indirect) implications of source dataset}}$

*At respective cutoff level.

of the implications. Within one set of MEC data, the total variance may be thought of as having two components: systematic variance and error (i.e., random) variance. The determination of the threshold cutoff defines what is systematic (i.e., will be represented in the HVM) and what is error (i.e., will not be represented in the HVM). The proportion of systematic variance to total variance may be thought of as a parallel measure for the *quasi*-coefficient of determination (r^2_{q}), of which the square root yields an estimate of *quasi*-reliability (r_q).

In terms of comparing the *quasi*-validity, or v_q, of two MEC solutions, the best estimate of a "true" set of connections is considered to be the set of common connections between two solutions (which are determined by respective threshold cutoff levels) (Kerlinger 1986, p. 427).[6] The error component for each solution is the number of unique implications represented from the connections that are different from the overlapping, or common, connections. In terms of MEC summary HVMs, this error component can be defined as the percentage of the total number of implications for an HVM solution, at a given cutoff value, that are not represented in the connections but are common between the two methods. It is important to note that the relative amount of overlap for each method may not be equal, as the magnitude of implications for the common connections may differ. Using this conceptual framework, it follows that the measure of *quasi*-validity (v_q) is the proportion of common implications, with respect to all implications computed from the MEC data for a given data set. Given that in any specific MEC data set, both the number of implications and the number of common relations may vary, the computation of the proportion measure of *quasi*-validity (v_q) will reflect both of these measures.

Thus, the issue of comparing two MEC solutions involves utilizing the principles of Ordinal Test Theory (Cliff and Keats 2003, ch. 6) to estimate at specified threshold (cutoff) levels: (1) the consistent implications contained in the solution as a basis to compute *quasi*-reliability (r_q, the proportion of implications accounted for in a given solution corresponds to the coefficient of determination [r^2_{q}], which means its square root is an estimate of reliability), and (2) the common variance between the two solutions as the basis to construct a measure of *quasi*-validity (v_q, which is the proportion of common implications to the total number of implications present in a data set).

Appendix B provides a simple example that illustrates the calculation of these quality metrics. It is our hope that reporting these metrics will become standard practice for all laddering research. In the following section, we report on a study from which the values of these metrics are then used as the basis of our analysis that compares two laddering approaches.

Study

Our study entails a comparative analysis of two approaches for collecting laddering data: a traditional face-to-face interviewing approach and an online one-on-one interviewing approach. The traditional face-to-face interviews were conducted by a leading professional research firm recognized for its expertise in laddering research. Our objective in selecting this firm was to obtain the highest-quality face-to-face laddering interviews commercially available. Moreover, given the comparative nature of our study, we requested that the firm use its top three interviewers to add an additional degree of rigor to this approach. The online approach was deemed to have significant potential to reduce the cost barriers associated with laddering research. These cost reductions were expected to be achieved in terms of increased interviewer productivity, ease of coding (i.e., the data are already in written form), and less time required to complete the project overall. In this section, we review and compare traditional face-to-face interviewing to online interviewing following the procedural operations detailed earlier; however, we incorporate a new "quality assessment" operation where we report the quality metrics proposed earlier (see Table 6.2). These metrics serve as the bases for our comparison of the two interviewing methods. Table 6.3 summarizes the key research dimensions for both approaches for each of the laddering operations.

Research/Questionnaire Design and Sample Specification

As the comparative analysis begins with the third laddering operation, interviewing, it was imperative that the first two operations, research/questionnaire design and sample specification, were held constant across both approaches. Our description of these two operations follows the four framing questions introduced earlier.

The domain of our investigation was the 2004 U.S. presidential election, which was chosen for its substantive interest and topical importance.[7] Although the interviews were conducted between late May and late July 2004, prior to the nominating conventions, state primary elections had confirmed in mid-March 2004 that the major party candidates for the November 2004 election would be the incumbent President George W. Bush for the Republicans and Senator John Kerry for the Democrats.

Given the substantive topic, the key decision we sought to understand was the voting intention of likely voters (i.e., *relevant consumers*), defined as individuals who were both eligible and intended to vote in the 2004 election, and who had voted in the prior presidential election in November 2000 (i.e., *particular behaviors or actions*). We should note that these same criteria are used to identify likely voters in many political surveys. As voting behavior is often associated with socio-demographic factors, we balanced our samples across three additional criteria (i.e., *social and physical contexts*): (1) vote intention, as measured on a seven-point scale from Definitely Bush to Definitely Kerry, with Undecided at the mid-point;[8] (2) age, classified as under 40 and over 40; and (3) gender. Finally, we identified the two major party candidates—incumbent President George W. Bush and Senator John Kerry—that likely voters would consider in making their voting decisions (i.e., *choice alternatives*).

The questionnaire asked two types of laddering questions appropriate to political research: *Top of Mind* and *Disequity*. The *top of mind* question asks the respondent, "*What is the very first thing that comes to your mind when you think of the (political party) candidate, (name)?*" and continues with a follow-up question, "*Is your top of mind comment about (candidate name) a positive or a negative to you?*" With the response to the first question and valence from the second question, the initial laddering probe is "*Why is that (i.e., top of mind mention) is (a positive or a negative) to you?*" Then standard laddering interviewing protocol comes into play. The top of mind laddering was done for both candidates, resulting in two ladders from each respondent.

The second type of laddering question first focused on obtaining a *disequity* of the candidate. Stated another way, this question tries to identify the primary barrier that keeps the voter from having a stronger intention to vote for that candidate. Specifically, the *disequity* question asked, "*Think carefully for a minute. What is the single most important thing—a position on a specific issue, or a leadership trait—that if changed about (candidate name), would make you more likely to vote for him?*" The rationale for this bipolar, "on the margin" question is that knowing what would move voters in a desired direction (e.g., from a vote intention of *most likely* to *definitely*) could be used in the strategy development process (cf. Kalwani and Silk 1982). For obvious reasons, this "barrier" question was not asked for the candidate for whom voters indicated a "definite" vote intention. These voters did, however, answer the disequity question for the candidate for whom they did not intend to vote. Thus, three ladders (i.e., two top of mind and one disequity) were obtained from voters who expressed a "definite" vote intention, whereas four ladders (i.e., two top of mind and two disequity) were obtained from respondents with a "less-than-definite" vote intention.[9]

Table 6.3

Comparison of Laddering Approach

Laddering Operation	Face-to-Face	Online
1. Research Design		
Domain	2004 presidential election	2004 presidential election
Laddering questions	3–4	3–4
2. Sample Specification		
Relevant consumers	Likely voters	Likely voters
Particular behavior or actions	Voted in 2000 election	Voted in 2000 election
Social and physical contexts		
Sex		
Female/Male	36/36	36/36
Age		
Under 40/Over 40	33/39	33/39
Vote intention		
Definitely Bush	10	10
Most Likely Bush	10	10
Leaning Bush	10	10
Undecided	11	11
Leaning Kerry	10	10
Most Likely Kerry	11	10
Definitely Kerry	10	11
Choice Alternatives	Bush vs. Kerry	Bush vs. Kerry
3. Interviewing		
No. of participants/sampling frame	72/major western city	72/national online panel
Interviewers	3 professional	5 part-time
Procedure		
Location	Research facility	Home or office
Communication mode	Talk	Talk and type
Average length of interview	35 minutes	36 minutes
Perceived accuracy	3.90/5	4.49/5**
4. Coding		
Number of codes	37	37
Number of ladder elements coded	1,256	1,133
Total ladders	267	261
Average ladder length	4.70**	4.34
Total implications	2,410	1,937
5. Quality Metrics		
Coding quality (C_Q)	84.4%	85.5%
Reliability index (I_r)	.916	.922
Ladder quality (L_Q)	65.9%	97.3%**
Quasi-coeff. of determination† (r_q^2)	.625*	.498
Quasi-Reliability† (r_q)	.788*	.701
Quasi-Validity† (v_q)	.467	.746**

*$p < .05$; **$p < .01$; †average across four ladder types for each of three threshold cutoff levels (Appendix D).

Interviewing

Participants

Respondents for the face-to-face laddering approach were pre-recruited at one central research facility located in a major western city. Respondents were screened according to the sample specification criteria described above, and seventy-two likely voters matching these criteria were selected to participate. All respondents completing the traditional laddering interview received a nominal monetary award for their participation.

Respondents for the online laddering approach were members of a national online discontinuous consumer access panel. Members were invited to register for a unique survey about the upcoming 2004 presidential election that was being conducted by a midwestern university. The invitation stated, "*What makes this survey unique (and fun!) is that this study uses a computerized interactive "chat" format so that we can better understand your thoughts and opinions.*" Panel members were also informed that only 72 participants would be selected and that registrants MUST: (1) have voted in the 2000 presidential election, (2) have Internet access, and (3) be able to hear sound on their computer, as their interviewer would be speaking to them live via the Internet. All registrants received an entry into the panel's quarterly drawing for a cash prize, and the registrants who were selected received a nominal monetary award as an additional incentive. The online registration form asked the members for their email address, age, gender, whether they had voted in 2000, and their vote intention, as measured on a seven-point scale, from Definitely Bush to Definitely Kerry. Registrants were also asked to list three convenient times to schedule an interview. Participants were selected according to the criteria described above and were contacted by the researchers via email, informing them of their interview time, their interviewer's name, and the web address of the interview.[10] Participants were instructed to simply click on the web address link in the email at the time of the interview and their interviewer would meet them there.

Interviewers

Three professional full-time laddering interviewers were used for the face-to-face approach. All interviewers were extensively trained and had an average of five years of professional experience conducting laddering interviews to elicit MECs.

The online approached used five part-time interviewers, all of whom were college educated and had some interviewing experience, but had minimal experience conducting laddering interviews. Online interviewers received approximately four days of training that involved (a) readings and discussions explaining means-end theory; (b) learning the laddering techniques used to probe for higher levels of abstraction; (c) rehearsing the interviewing techniques, including identifying respondent comments as to their level of abstraction; and (d) conducting practice interviews using the computer interface, which required them to determine the level of abstraction of each response the respondent made.

Procedure

Interviews using the traditional face-to-face approach were conducted in person at the research facility. Upon arriving at the facility, participants were greeted by their interviewer, who then read a brief introduction to the survey. The interviewers recorded the start and ending time of each interview, which averaged approximately 35 minutes. To control for possible order effects, the interviewers were instructed to rotate the order of presentation of the candidates. In addition, respondents were shown a card for questions containing the list of possible response alternatives. At the conclusion of the interview, respondents were asked to evaluate the accuracy of their answers to the laddering questions. Respondents rated the accuracy of their answers to be quite high (on a five point scale, with 5 = Perfectly accurate, 4 = Very accurate, 3 = Fairly accurate; 2 = Somewhat accurate; 1 = Not accurate; mean = 3.90). Traditional interviewers recorded the respondents' ladders by typing them into a computer using a common word-processing software program while they were interviewing, and the 72 transcripts were sent to the authors. The respondents' verbatim comments to the open-ended laddering questions were listed in the order in which they were elicited. That is, the ladder elements (i.e., verbatim comments) were not necessarily in hierarchical order.

Online interviews were conducted live via the Internet using a specially designed presentation-and-response format for laddering. In this approach, the respondent is queried with a combination of oral (voice-over-net) and written questioning (video text) and is asked to respond via an online "point-and-click" (i.e., mouse) mode for questions using a standardized fixed-choice response format, and a "chat" (i.e., typing) mode for open-ended questions. It should be noted that the respondent could "chat" with the interviewer at any time during the interview by typing in the respondent's box on the screen. In our online study, the software recorded the starting and ending time of each session, and the sessions averaged approximately 50 minutes. Online sessions included both an introductory practice segment lasting approximately 14 minutes and a focal interview on the election lasting approximately 36 minutes. In addition, to control for possible order effects, the software randomized the order of presentation of the candidates and the fixed-choice response options. Ladder elements were reviewed by the interviewer in real-time. That is, as the respondent typed in his or her reply, an interviewer made a determination as to the "level" (i.e., attribute, consequence, or value) of each response. The interviewer would then enter the appropriate summary into the matching "*rung*" of the ladder display on the interviewer's screen. Next, the interviewer would proceed to probe the respondent by voice for the next level. This process continued until the interviewer determined that a complete ladder (consisting of responses at all four levels) had been achieved. Once the interview was completed, the interviewer would visually display the complete ladder on the respondent's screen, while providing a verbal summary of the complete ladder to the respondent. The respondent was then asked to confirm the accuracy of his or her ladder. If the respondent indicated a change was needed, the interviewer would modify the ladder until the respondent confirmed its integrity. At the conclusion of the interview, the respondent evaluated the accuracy of his or her answers to the laddering questions. Respondents rated the accuracy of their answers on the same five-point accuracy scale described above (mean = 4.49).

Comparison

We begin by examining the commonalities and contrasts related to the third laddering operation, interviewing. The first notable contrast is the sampling frame. The face-to-face approach drew its sample from a single geographic location, namely, a major western city. In contrast, the online approach selected its participants from a nationwide online panel. As ter Hofstede et al. (1998) and others have remarked, one of the major limitations of laddering has been the impracticality of obtaining geographically representative samples. Clearly, the online approach makes reaching geographically diverse samples a possibility.

The second notable difference concerns interviewer experience/training. The face-to-face method used three professional full-time interviewers with an average of five years of laddering experience, whereas the online approach used five newly trained part-time interviewers with minimal laddering experience. As discussed at length, laddering interviews require great skill and have traditionally required highly trained interviewers. The online approach hastened mastery of the needed skills by providing the inexperienced interviewers with visual self-checks and procedures to evaluate the quality and completeness of each ladder obtained. It is important to note that the online interviewers were trained to use proper interviewing techniques and to continue questioning the respondent *until* a complete ladder was obtained, and they knew they would be evaluated in this manner. Although the face-to-face (professional) interviewers were more experienced, it is possible that they were also more complacent. Our experience suggests that interviewers rarely use "chutes," or reverse laddering. That is, if the respondent skips the functional consequence level and reports a psychosocial consequence (e.g., feels secure), the interviewer does not go "back down" to make sure he/she elicits a response at the lower level (e.g., "What is it about brand X that makes you feel secure?"). The online interviewing system ensures that the interviewer is cognizant of this goal for each interview.

A third difference concerns interview location. The face-to-face approach restricted interviews to one centrally located research facility. In contrast, the online approach offered greater time and place flexibility to both the interviewer and the respondent, in that the interview could be conducted at any location with Internet access. Thus, the online approach could significantly benefit professional researchers who consider the time and cost of interviewing, along with finding qualified interviewers, to be among the greatest obstacles to conducting laddering research. Although home Internet access is certainly growing, its requirement is an obvious limitation for researchers needing to reach respondents at lower socioeconomic levels who many not have Internet access, or respondents for whom literacy is an issue. For these respondents, the "talking" face-to-face approach is undoubtedly more pragmatic than the "typing" online approach. While these communication mode differences may indeed present hurdles for some special populations, it is also likely that many researchers would readily warrant the additional cognition required of respondents asked to type rather than simply speak their replies. When the prospect of more thoughtful responses is coupled with the additional time- and cost-savings resulting from respondents keying their own data, the online approach appears particularly advantageous.

It is important to note that unlike previous online "hard" laddering approaches (e.g., Russell et al. 2004a, 2004b), we have demonstrated that online laddering does not necessarily need to be "hard." Our online approach did not require that our respondents produce ladder elements in a sequential ordered progression from low to high, nor did it overly restrict the respondent's ability to answer (Grunert and Grunert 1995). The elaborations, twists and turns, and jumping back and forth between different levels of abstraction were possible and did occur. That is, although respondents were required to type their answers, they were not restricted in their responses. This is not to imply that the online approach is as "soft" as the face-to-face approach, but rather to note that much of the fluidity of the interviewing protocol in the online one-on-one approach was maintained.

Although the average length of an overall interview differed by about 15 minutes, the difference in length of the focal interview itself was not significant ($t = -.003$, n.s.). Nearly all of the additional 15 minutes for the online approach can be relegated to a practice segment. This practice segment not only allowed the respondents to familiarize themselves with the online controls, but also enabled the interviewer to ensure system compatibility for the focal interview. The final difference noted in this section relates to the respondents' perceived accuracy of their answers to the laddering questions, as measured on the five-point scale presented above that ranged from "Not accurate" to "Perfectly accurate." Here, we see a statistically significant difference with face-to-face respondents reporting lower accuracy (mean = 3.90) than the online respondents (mean = 4.49; $t = -6.13$, $p < .01$, two-tailed). It is likely that this difference is attributable to the online interviewing procedures that had respondents both visually and aurally review their ladders and confirm their accuracy to the interviewers.

Coding

After an initial reading and classification of the open-ended (i.e., verbatim) responses to the two types of laddering questions (i.e., top of mind and disequity) for both the face-to-face and online studies, a coding scheme was developed to categorize the responses of each question. The thirty-seven coding categories and their distributions within each of the four levels of abstraction are listed in Appendix C. Two coders, one blind to the experimental conditions, independently assigned a code to each ladder element. Coding time averaged approximately 3.5 minutes per ladder. The coding operation of the data from the traditional approach yielded a total of 267 ladders (i.e., MECs) consisting of 1,256 ladder elements with an average ladder length of 4.70 elements. In contrast, the online approach yielded 261 ladders consisting of 1,133 ladder elements with an average ladder length of 4.34 elements. The difference in ladder length between the two approaches was statistically significant ($t = 5.98$, $p < .01$, two-tailed) indicating that the ladders elicited with the face-to-face approach were significantly longer.

Quality Assessment

Coding Quality (C_Q)

Coding quality (C_Q) reflects the inter-coder agreement on matching each of the ladder elements (face-to-face = 1,256; online = 1,133) with the same code from a code list that numbered 37 codes. Discrepancies were resolved by a judge. C_Q for both face-to-face (84.4 percent) and online (85.5 percent) was good and did not differ significantly ($\chi^2 = .59$, $df = 1$, n.s.). Inter-coder reliability was also strong, as indicated by the reliability index (I_r), which equaled .92 for each data set. To better understand the coding discrepancies above, we calculated coder agreement on matching ladder elements with the same level of abstraction. The high agreement levels for each approach (face-to-face $C_Q = 96.1\%$, $I_r = .974$; online $C_Q = 99.5$ percent; $I_r = .997$) suggests that discrepancies tended to occur within the same level of abstraction.

Ladder Quality (L_Q)

Ladder quality (L_Q) reflects the percentage of total ladders that have at least one ladder element at each of the four levels of abstraction. Of the 267 ladders elicited from the face-to-face approach only 176 were complete ($L_Q = 65.9$ percent). In contrast, 254 of the 261 ladders from the online approach were complete ($L_Q = 97.3$ percent). This difference in ladder quality (L_Q) between the two approaches is statistically significant ($\chi^2 = 86.1$, $df = 1$, $p < .01$). The success of the online approach may be attributed to requiring the interviewer to classify the level of each response during the interview, thereby focusing him or her on obtaining all four levels. Given that a complete ladder generally consists of four (minimum) to six elements, it is interesting to note that, despite having longer ladders, over one-third of the ladders yielded by the face-to-face approach were incomplete.

The significant difference in the ladder quality (L_Q) dramatizes the importance of reporting the metric of complete ladders for all means-end research. The laddering methodology is based upon obtaining complete ladders, and quality shortfalls may make the subsequent analyses, upon which strategy is based, questionable. The online approach clearly demonstrated that obtaining high-quality ladders is possible and should serve as an important benchmark for laddering researchers. It is worth noting that this finding is not an indicator of the benefits of online interviewing per se, but rather a demonstration that when quality metrics such as L_Q exist, they can be applied to obtain better results.

Quasi-Reliability (r_q)

To assess ladder reliability, we calculated *quasi*-coefficients of determination (r^2_{q}) for each of the four ladder types, at three separate threshold cutoff values (i.e., 3, 4, 5). Each of the resulting *quasi*-coefficients of determination (r^2_{q}) signifies the proportion of total variance (i.e., implications) considered to be systematic (see Appendix D).[11] The average of the twelve *quasi*-coefficients of determination (r^2_{q}) calculated from all (i.e., direct + indirect) significant implications suggest that for the face-to-face approach 62.8 percent of total variance may be considered systematic and would be represented in the resulting HVM, whereas 37.2 percent may be considered error variance and would not be represented in the HVM. For the online approach, 49.8 percent of the variance may be considered systematic while 50.2 percent may be considered error. The resulting twelve *quasi*-reliabilities (r_q) are based on the significant "common" connections within the ladders (face-to-face = 267; online = 261). The average of these twelve *quasi*-reliabilities (r_q) is .788 for the face-to-face approach and .701 for the online approach, both of which can be considered acceptable based on conventional standards.

Although the key reliability metric (r_q) to examine should be based on *all* (direct + indirect) significant pathways, as a point of reference we also calculated the *quasi*-coefficients of determination (r^2_{q}) and the resulting *quasi*-reliabilities (r_q) for significant direct connections only (see Appendix D). For the face-to-face approach the average of these twelve *quasi*-reliabilities (r_q) is .719, which is significantly lower than the average *quasi*-reliability for all significant connections ($t = -2.28$, $p < .05$, one-tailed). For the online approach the average of the *quasi*-reliabilities (r_q) for the direct connections is .665, which is also lower than the *quasi*-reliability for all significant connections, but this difference is not statistically significant ($t = -.931$, n.s.). Had the *quasi*-reliability of the significant direct implications been higher for either approach, it would mean a different interpretation of the resulting HVM.

The *quasi*-reliability (r_q) of both approaches may be judged as satisfactory; however, the average r_q of the face-to-face approach (.788) was deemed to be significantly higher than for the online approach (.701; $t = 2.51$, $p < .05$, two-tailed). The key driver of *quasi*-reliability is the commonality of code-pairs. A post-hoc analysis of the code distribution found significant differences between the two approaches within the psychosocial and values levels (see Appendix C). It is likely that the differences within each level are attributable to the geographic differences within the samples. Whereas the online approach used a nationwide panel, the face-to-face approach drew its sample entirely from one western city with a strong religious orientation.

Directly related to this finding is the *quasi*-coefficient of determination (r^2_{q}), which suggests that the face-to-face approach had a significantly higher proportion (.625) of its average total significant implications (i.e., variance) considered systematic than did the online approach (.498; $t = 2.53$, $p < .05$, two-tailed). In interpreting these metrics, it is useful to consider the nature of political data, which can differ considerably from data collected within the consumer domain. Generally, there are far fewer meaningful attribute distinctions for a given consumer product (e.g., ketchup) than there are for a particular presidential candidate who may stand for or embody any number of policy issues or leadership traits (e.g., Bahner and Fiedler 1985). Thus, capturing approximately 60 percent or better of the systematic variance with a threshold cutoff of 3 is quite reasonable for both approaches.

Quasi-Validity (v_q)

The *quasi*-validity metric (v_q) reflects the "true" structure obtained from two different methods. That is, v_q is the proportion of the significant "common" connections (i.e., implications) at a given threshold level that exist for both approaches. Thus, v_q is the most important quality metric, as it is these "common" connections that lead to the same interpretation across methods (i.e., HVMs in the "data summary" operation and strategy development in the final "interpretation" operation). Therefore, v_q indicates how accurate a given method is in producing consistent results. To calculate v_q we contrasted the two approaches to determine the common implications (see Appendix D). For the face-to-face approach, the average of the twelve *quasi*-validity metrics (v_q) for all significant connections (i.e., direct and indirect) is .467. For comparison, we also calculated v_q for the significant direct connections only. As expected, the average of these twelve *quasi*-validity metrics (v_q) is .453, which is lower than the average for all significant connections, but this difference was not statistically significant ($t = -.417$, n.s.). For the online approach, the average of the twelve *quasi*-validity metrics (v_q) for all significant connections is .746. For comparison, the average of the direct connections only is .632, which is significantly lower than the average for all direct and indirect connections ($t = -2.74$, $p < .01$, one-tailed).

Although reliability is a measure of internal consistency, it is validity that best reflects the "true" structure, as v_q is the shared systematic variance across two methods. Here, we see that the online approach reports an average v_q of .746, which is significantly higher than the .467 reported for the face-to-face approach ($t = -8.38$, $p < .01$, two-tailed).

As we reported earlier, there as been some debate about differences between "hard" and "soft" laddering approaches (e.g., Gurnert and Grunert 1995; Phillips and Reynolds 2009; Russell 2004a, 2004b). Several authors have suggested that comparisons be made (see Russell 2004a for a review), but the benchmarks and quality metrics needed for comparison have been lacking. Our metrics should allow for methodologically rigorous assessments of different laddering approaches. As we note at the conclusion of Appendix B, a simple visual comparison of HVMs may be misleading. Therefore, these metrics can serve as the needed benchmarks by which comparisons can be made and best practices developed.

Data Summary and Interpretation

A summary of the quality metrics discussed above is presented in Table 6.3. In interpreting these metrics, it is important to consider the final two laddering research operations: data summary (constructing the HVM) and interpretation (developing strategy). Marketing researchers use laddering to gain insight into consumer decision making. These insights are then used to develop customer-oriented strategies. As the final two procedures are entirely dependent on the quality of the data resulting from the first four operations (i.e., design, sample, interviewing, and coding), our motivation in introducing the quality assessment procedures and metrics was to enable researchers to ensure the reliability of the strategic insights captured from the laddering technique. In addition, we used these proposed metrics to compare the validity of two laddering methods. The overall results of our comparison clearly demonstrate the effectiveness of the online approach to yield quality, reliable, and valid ladders.

Concluding Remarks

Laddering is an important conceptual paradigm that has not reached its potential. Laddering research requires an interviewing depth that cannot be readily captured through other techniques and that yields a level of understanding unequaled by standardized questionnaires. The results from exit polls conducted after the 2004 U.S. presidential election demonstrate that, as we seek to better understand the drivers of behavior and decision making, we need to be wary of findings that may be artifacts associated with the method of inquiry.

Vargo and Lusch's (2004) paradigm of service-centered exchange views tangible goods as the provision of satisfaction for higher-order needs (cf. Gutman 1982) and calls for customers to be co-producers in the value creation process. The laddering methodology is uniquely positioned to help marketers achieve this aim by better understanding how customer values drive choice—and that ultimately determines what customers value (cf. Woodruff 1997). Yet, the current methodology will need to be modified (cf., Duhachek, Coughlan, and Iacobucci 2005; Rust 2004; Shugan 2004) as laddering researchers merit greater confidence in the validity of their findings. Our hope is that the quality metrics and procedures proposed here become standard, and that these metrics are reported by all laddering researchers—academics and practitioners alike.

Despite the value that marketing researchers place on means-end research and the deep insight into consumer decision making that the laddering technique reveals for their clients, the technique has presented researchers with noteworthy obstacles. Chief among these have been the time and costs of interviewing and coding, along with finding qualified interviewers. These obstacles alone have resulted in researchers either forgoing the use of laddering for projects to which it would be well suited, or taking procedural shortcuts that could hinder data quality and interpretation. As the online approach demonstrated, technology can ease the time and cost burden, while also yielding high-quality, reliable, and valid results.

Acknowledgments

Both authors contributed equally to this research project; the order of authors' names was randomly determined.

We gratefully acknowledge the helpful guidance provided by the editor and the reviewers. We also thank Tom Greene for his technical assistance, and Joe Urbany and Mark Parry for their helpful comments on earlier drafts of this chapter. This project was supported by a grant to Joan M. Phillips from the Mendoza College of Business, University of Notre Dame, and was conducted while Thomas J. Reynolds was Visiting Professor of Marketing and Joan M. Phillips was Assistant Professor of Marketing at the University of Notre Dame.

Notes

1. This ad, also known as "The Bear in the Woods," can be viewed at www.livingroomcandidate.org, a website maintained by the American Museum of the Moving Image.

2. See also Reynolds and Gutman (1988) and Reynolds and Olson (2001) for a thorough technical description of these six laddering operations.

3. The importance of this grouping decision is well illustrated in the example above that compared the HVM obtained from Reagan's supporters to the HVM obtained from Mondale's. The studies presented in Table 6.1, particularly those in the cross-cultural category, provide many additional examples.

4. Such failure to adhere to rigorous "textbook procedures" likely occurs broadly (beyond laddering) across many analytical techniques.

5. HVMs may become too complex to interpret if every implication is included. Therefore, researchers are advised to select a cutoff level that would explain about 70 percent of the total (direct and indirect) implications accounted for.

6. In larger samples (e.g., ~120 ladders) a split-half analysis could be used to estimate validity.

7. The study of voting behavior and political campaigns has long been considered to be within the domain of marketing (e.g., Kotler and Levy 1969). For examples of marketing's contributions within this substantive domain see: Chapman and Palda 1983, 1984; Crosby, Gill, and Taylor 1981; Heeler and Ray 1972; Klein and Ahluwalia 2005; Morwitz and Pluzinski 1996; Nakanishi, Cooper, and Kassarjian 1974; Newman and Sheth 1985; Phillips, Urbany, and Reynolds 2008; Rothschild 1978; Simmons, Bickart, and Lynch 1993.

8. Actual scale values were: Definitely Bush, Most Likely Bush, Leaning Toward Bush, Undecided, Leaning Toward Kerry, Most Likely Kerry, Definitely Kerry

9. To simulate a complete laddering interview session, a range of additional questions (27 total) were incorporated in the design, including demographic, quantitative scales, and open-ended questions.

10. As this was part of a slightly larger study (n ~ 100) on political strategy, by design some cells (e.g., Definitely Bush and Definitely Kerry) were oversampled and contained more respondents than needed. Therefore, respondents within these cells were randomly selected to closely match the sample distribution used in the face-to-face study.

11. By definition, the *quasi*-coefficient of determination will decrease as the threshold cutoff level increases since fewer implications are included in the solution.

References

Abelson, R. 1963. "Computer Simulation of 'Hot Cognition.'" In *Computer Simulation of Personality: Frontier of Psychological Theory*, ed. Silvan S. Tomkins and Samuel Messic. New York: John Wiley & Sons.

American Marketing Association. 2004. Marketing Services Directory. Accessed at www.marketingpower.com on July 15, 2004.

Aurifeille, Jacques-Marie, and Pierre Valette-Florence. 1995. "Determination of the Dominant Means-End Chains: A Constrained Clustering Approach." *International Journal of Research in Marketing* 12 (October), 267–78.

Babington, Charles, and Brian Faler. 2004. "Sliding Scale of Moral Values Is All in the Phrasing." *Washington Post*, November 14, A05.

Bagozzi, Richard P., and Pratibha A. Dabholkar. 1994. "Consumer Recycling Goals and Their Effect on Decisions to Recycle: A Means-End Chain Analysis." *Psychology & Marketing* 11 (July/August), 313–40.

———. 2000. "Discursive Psychology: An Alternative Conceptual Foundation to Means-End Chain Theory." *Psychology & Marketing* 7 (July), 535–86.

Bagozzi, Richard P., and Utpal Dohlakia. 1999. "Goal Setting and Goal Striving in Consumer Behavior." *Journal of Marketing* 63 (Special Issue), 19–32.

Bahner, Lesley, and John Fiedler. 1985. "The Application of MECCAS to Advertising Strategy Development for the 1984 Reagan-Bush Presidential Campaign." Paper presented at the American Marketing Association Winter Educators' Conference, Phoenix, AZ.

Baker, Susan, Keith E. Thompson, and Julia Engelken. 2004. "Mapping the Values Driving Organic Food Choice." *European Journal of Marketing* 38 (8), 995–1012.

Baker, Susan, Keith E. Thompson, and David Palmer-Barnes. 2002. "Crisis in the Meat Industry: A Values-based Approach to Communications Strategy." *Journal of Marketing Communications* 8, 19–30.

Bech-Larsen, Tino. 1996. "Danish Consumers' Attitudes to the Functional and Environmental Characteristics of Food Packaging." *Journal of Consumer Policy* 19 (3), 339–63.

Blattberg Robert C., Gary Getz, and Jacquelyn S. Thomas. 2001. *Customer Equity: Building and Managing Relationships as Valuable Assets.* Boston: Harvard Business School Press.

Botschen, Günther, and Andrea Hemetsberger. 1998. "Diagnosing Means-End Structures to Determine the Degree of Potential Marketing Program Standardization." *Journal of Business Research* 42 (2), 151–59.

Botschen, Günther, Eva M. Thelen, and Rik Pieters. 1999. "Using Means-End Structures for Benefit Segmentation." *European Journal of Marketing* 33 (1/2), 38–58.

Bradburn, Norman M., Seymour Sudman, and Brian Wansink. 2004. *Asking Questions: The Definitive Guide to Questionnaire Design—For Market Research, Political Polls, and Social and Health Questionnaires.* San Francisco: Jossey-Bass.

Bredahl, L. 1999. "Consumers' Cognitions with Regard to Genetically Modified Foods: Results of a Qualitative Study in Four Countries." *Appetite* 33 (3), 343–60.

Chapman, Randall G., and Kristian S. Palda. 1984. "Assessing the Influence of Campaign Expenditures on Voting Behavior with a Comprehensive Electoral Market Model." *Marketing Science* 3 (Summer), 207–26.

———. 1983. "Electoral Turnout in Rational Voting and Consumption Perspectives." *Journal of Consumer Research* 9 (March), 337–46.

Chiu, Chao-Min. 2005. "Applying Means-End Chain Theory to Eliciting System Requirements and Understanding Users Perceptual Orientations." *Information & Management* 42, 455–68.

Clarke, G., and L.W. Murray. 2001. "Frames of Reference in Financial Corporate Governance and Communications." *Corporate Governance* 1 (4), 20–27.

Cliff, Norman, and John A. Keats. 2003. *Ordinal Measurement in the Behavioral Sciences.* Mahwah, NJ: Lawrence Erlbaum.

Cohen, Joel, and Luk Warlop. 2001. "A Motivational Perspective on Means-End Chains." In *Understanding Consumer Decision Making: The Means-End Approach to Marketing and Advertising Strategy*, ed. Thomas J. Reynolds and Jerry C. Olson. Mahwah, NJ: Lawrence Erlbaum.

Coolen, Henny, and Joris Hoekstra. 2001. "Values as Determinants of Preferences for Housing Attributes." *Journal of Housing and the Built Environment* 16, 285–306.

Coombs, Clyde H. 1964. *A Theory of Data.* New York: John Wiley & Sons.

Cooper, Harris M.. 1998. *Synthesizing Research: A Guide for Literature Reviews.* Thousand Oaks, CA: Sage.

Crosby, Lawrence A., James D. Gill, and James R. Taylor. 1981. "Consumer/Voter Behavior in the Passage of the Michigan Container Law." *Journal of Marketing* 45 (Spring), 19–32.

Dabholkar, Pratibha A. 1994. "Incorporating Choice into an Attitudinal Framework: Analyzing Models of Mental Comparison Processes." *Journal of Consumer Research* 21 (June), 100–118.

Deeter-Schmelz, Dawn R., Karen Norman Kennedy, and Daniel J. Goebel. 2002. "Understanding Sales Manager Effectiveness: Linking Attributes to Sales Force Values." *Industrial Marketing Management* 31 (7), 617–26.

Deeter-Schmelz, Dawn R., and Jane Z. Sojka. 2004. "Wrestling with American Values: An Exploratory Investigation of World Wrestling Entertainment as a Product-based Subculture." *Journal of Consumer Behaviour* 4 (2), 132–43.

Dennett, Daniel C. 1987. *The Intentional Stance.* Cambridge, MA: MIT Press.

Devlin, Derek, Grete Birtwistle, and Norma Macedo. 2003. "Food Retail Positioning Strategy: A Means-End Chain Analysis." *British Food Journal* 105 (9), 653–70.

Dibley, Anne and Susan Baker. 2001. "Uncovering the Links Between Brand Choice and Personal Values Among Young British and Spanish Girls." *Journal of Consumer Behaviour* 1 (1), 77–93.

Duhachek, Adam, Anne T. Coughlan, and Dawn Iacobucci. 2005. "Results of the Standard Error of the Coefficient Alpha Index of Reliability." *Marketing Science* 24 (Spring), 294–301.

Flight, I, C.G. Russell, I. Blossfeld, and D.N. Cox. 2003. "From Sensory Attributes to Marketing Hooks: Using Laddering to Understand Consumer Perceptions of Red Meat." *Food Australia* 55 (9), 418–24.

Fotopoulos, Christos, Athanasios Krystallis, and Mitchell Ness. 2003. "Wine Produced by Organic Grapes in Greece: Using Means-End Chains Analysis to Reveal Organic Buyers' Purchasing Motives in Comparison to the Non-Buyers." *Food Quality and Preference* 14 (7), 549–66.

Gengler, Charles E. 1995. *LADDERMAP.* Camden, NJ: Means-End Software.

Gengler, Charles E., Daniel J. Howard, and Kyle Zolner. 1995. "A Personal Construct Analysis of Adaptive Selling and Sales Experience." *Psychology and Marketing* 12 (4), 287–304.

Gengler, Charles E., David B. Klenosky, and Michael S. Mulvey. 1995. "Improving the Graphic Representation of Means-End Results." *International Journal of Research in Marketing* 12 (October), 245–56.

Gengler, Charles E., Michael S. Mulvey, and Janet E. Oglethorpe. 1999. "A Means-End Analysis of Mothers' Infant Feeding Choices." *Journal of Public Policy and Marketing* 18 (2), 172–88.

Gengler, Charles E., and Thomas J. Reynolds. 1995. "Consumer Understanding and Advertising Strategy: Analysis and Strategic Translation of Laddering Data." *Journal of Advertising Research* 35 (July/August), 19–33.

Goldenberg, Marni A., David B. Klenosky, Joseph T. O'Leary, and Thomas J. Templin. 2000. "A Means-End Investigation of Ropes Course Experiences." *Journal of Leisure Research* 32 (2), 208–24.

Griffin, Abbie, and John R. Hauser. 1982. "The Voice of the Customer." *Marketing Science* 12 (Winter) 1–27.

Grunert Klaus G., and Tino Bech-Larsen. 2005. "Explaining Choice Option Attractiveness by Beliefs Elicited by the Laddering Method." *Journal of Economic Psychology* 26, 223–41.

Grunert, Klaus G., Suzanne C. Beckmann, and Elin Sørensen. 2001. "Means-End Chains and Laddering: An Inventory of Problems and an Agenda for Research." In *Understanding Consumer Decision Making: The Means-End Approach to Marketing and Advertising Strategy*, ed. Thomas J. Reynolds and Jerry C. Olson. Mahwah, NJ: Lawrence Erlbaum.

Grunert, Klaus G., and Suzanne C. Grunert. 1995. "Measuring Subjective Meaning Structures by the Laddering Method: Theoretical Considerations and Methodological Problems." *International Journal of Research in Marketing* 12 (October), 209–25.

Grunert, Klaus G., Liisa Lähteenmäki, Niels Asger Nielsen, Jacob B. Poulsen, Oydis Ueland, and Annika Astrom. 2001. "Consumer Perceptions of Food Products Involving Genetic Modification: Results from a Qualitative Study in Four Nordic Countries." *Food Quality and Preference* 12 (8), 527–42.

Grunert, Klaus G., and Carlotta Valli. 2001. "Designer-made Meat and Dairy Products: Consumer-led Product Development." *Livestock Production Science* 72 (1–2), 83–98.

Guenzi, Paolo, and Gabriele Troilo. 2006. "Developing Marketing Capabilities for Customer Value Creation Through Marketing-Sales Integration." *Industrial Marketing Management* 35, 974–88.

Guiltinan, Joseph P. 2002. "Choice and Variety in Antitrust Law: A Marketing Perspective." *Journal of Public Policy & Marketing* 21 (Fall), 260–68.

Gutman, Jonathan. 1982. "A Means-End Chain Model Based on Consumer Categorization Processes." *Journal of Marketing* 46 (Spring), 60–72.

———. 1990. "Adding Meaning to Values by Directly Assessing Value–Benefit Relationships." *Journal of Business Research* 20, 153–60.

———. 1997. "Means-End Chains as Goal Hierarchies." *Psychology and Marketing* 14 (September), 545–60.

Gutman, Jonathan, and George Miaoulis. 2003. "Communicating a Quality Position in Service Delivery: An Application in Higher Education." *Managing Service Quality* 13 (2), 105–11.

Gutman, Jonathan, and Thomas J. Reynolds. 1979. "An Investigation of the Levels of Cognitive Abstraction Utilized by Consumers in Product Differentiation." In *Attitude Research Under the Sun*, ed. J. Eighmey. Chicago: American Marketing Association.

Hakkio, Satu, and Pirjo Laaksonen. 1998. "Relationships in Marketing Channels: Examining Communication Abilities Through Cognitive Structures." *Psychology and Marketing* 15 (3), 215–40.

Haras, Kathy, Camille J. Bunting, and Peter A. Witt. 2006. "Meaningful Involvement Opportunities in Ropes Course Programs." *Journal of Leisure Research* 38 (3), 339–62.

Heeler, Roger M., and Michael L. Ray. 1972. "Measure Validation in Marketing." *Journal of Marketing Research* 9 (November), 361–70.

Huber, Frank, Suzanne C. Beckmann, and Andreas Herrmann. 2004. "Means-End Analysis: Does the Affective State Influence Information Processing Style?" *Psychology & Marketing* 21 (9), 715–37.

Jaeger, Sara R., and Halliday J.H. MacFie. 2001. "The Effect of Advertising Format and Means-End Information on Consumer Expectations for Apples." *Food Quality and Preference* 12 (3), 189–205.

Jain, Arvind K., and Annamma Joy. 1997. "Money Matters: An Exploratory Study of the Socio-Cultural Context of Consumption, Saving, and Investment Patterns." *Journal of Economic Psychology* 18 (6), 649–75.

Kalwani, Manohar U., and Alvin J. Silk. 1982. "On the Reliability and Predictive Validity of Purchase Intention Measures." *Marketing Science* 1 (Summer), 243–86.

Kamakura, Wagner A., and Thomas P. Novak. 1992. "Value-System Segmentation: Exploring the Meaning of LOV." *Journal of Consumer Research* 119–32.

Kassarjian, Harold. 1977. "Content Analysis in Consumer Research." *Journal of Consumer Research* 4 (June), 8–18.

Kelly, George A. 1955. *The Psychology of Personal Construct*, New York: W.W. Norton.

Kerlinger, Fred N. 1986. *Foundations of Behavioral Research.* 3rd ed. New York: Holt, Rinehart and Winston.

Klein, Jill G., and Rohini Ahluwalia. 2005. "Negativity in the Evaluation of Political Candidates." *Journal of Marketing* 69 (January), 131–42.

Klenosky, David B. 2002. "The 'Pull' of Tourism Destinations: A Means-End Investigation." *Journal of Travel Research* 40 (May), 385–95.

Klenosky, David B., Charles E. Gengler, and Michael S. Mulvey. 1993. "Understanding the Factors Influencing Ski Destination Choice—A Means-End Analytic Approach." *Journal of Leisure Research* 25 (4), 362–79.

Klenosky, David B., Thomas J. Templin, and Josh A. Troutman. 2001. "Recruiting Student Athletes: A Means-End Investigation of School-Choice Decision Making." *Journal of Sport Management* 15 (2), 95–106.

Kotler, Philip, and Sidney J. Levy. 1969. "Broadening the Concept of Marketing." *Journal of Marketing* 33 (January), 10–15.

Lageat, Thierry, Sandor Czellar, and Gilles Laurent. 2003. "Engineering Hedonic Attributes to Generate Perceptions of Luxury: Consumer Perception of an Everyday Sound." *Marketing Letters* 14 (2), 97–109.

Langerak, Fred, Ed Peelen, and Ed Nijssen. 1999. "A Laddering Approach to the Use of Methods and Techniques to Reduce the Cycle Time of New-to-the-Firm Products." *Journal of Product Innovation Management* 16 (2), 173–82.

Lastovicka, John L. 1995. "LADDERMAP—VERSION 4.0." *Journal of Marketing Research* 32 (November), 494–96.

Laukkanen, Tommi. 2006. "Customer-Perceived Value of E-Financial Services: A Means-End Approach." *International Journal of Electronic Finance* 1 (1), 5–17.

Laukkanen, Tommi, and Jari Lauronen. 2005. "Consumer Value Creation in Mobile Banking Services." *International Journal of Mobile Communications* 3 (4), 325–38.

Leppard, P., C.G. Russell, and D.N. Cox. 2004. "Improving Means-End-Chain Studies by Using a Ranking Method to Construct Hierarchical Value Maps." *Food Quality and Preference* 15, 489–97.

Lin, Chin-Feng. 2003. "Quality-Delivery System: A Conceptual Framework of Attribute Level-Value Linkages." *Total Quality Management and Business Excellence* 14 (10), 1079–92.

Ling, Chin-Feng, and Ming-Yih Yeh. 2000. "Means-End Chains and Cluster Analysis: An Integrated Approach to Improving Marketing Strategy." *Journal of Targeting, Measurement, and Analysis for Marketing* 9 (1), 20–35.

Makatouni, Aikaterini. 2002. "What Motivates Consumers to Buy Organic Food in the UK: Results from a Qualitative Study." *British Food Journal* 104 (3–5), 345–52.

Manyiwa, Simon, and Ian Crawford. 2002. "Determining Linkages Between Consumer Choices in a Social Context and the Consumer's Values: A Means-End Approach." *Journal of Consumer Behaviour* 2 (1), 54–70.

Menand, Louis. 2004. "Permanent Fatal Errors: Did the Voters Send a Message?" *New Yorker* 80 (38), 54–60.

Miles, Susan, and Lynn J. Frewer. 2001. "Investigating Specific Concerns About Different Food Hazards." *Food Quality and Preference* 12 (1), 47–61.

Mitchell, Vincent-Wayne, and Greg Harris. 2005. "The Importance of Consumers' Perceived Risk in Retail Strategy." *European Journal of Marketing* 39 (7/8), 821–37.

Morandin, Gabriele, Massimo Bergami, and Richard P. Bagozzi. 2006. "The Hierarchical Cognitive Structure of Entrepreneur Motivation Toward Private Equity Financing." *Venture Capital* 8 (3), 253–71.

Mort, Gillian Sullivan, and Trista Rose. 2004. "The Effects of Product Type on Value Linkages in the Means-End Chain: Implications for Theory and Method." *Journal of Consumer Behaviour* 3 (3), 221–34

Morwitz, Vicki G., and Carol Pluzinski. 1996. "Do Polls Reflect Opinions or Do Opinions Reflect Polls? The Impact of Political Polling on Voters' Expectations, Preferences, and Behavior." *Journal of Consumer Research* 23 (June), 53–67.

Mulvey, Michael S., Jerry C. Olson, Richard L. Celsi, and Beth A. Walker. 1994. "Exploring the Relationships Between Means-End Knowledge and Involvement." *Advances in Consumer Research* 21, 51–57.

Nakanishi, Masao, Lee G. Cooper, and Harold H. Kassarjian. 1974. "Voting for a Political Candidate Under Conditions of Minimal Information." *Journal of Consumer Research* 1 (September), 36–43.

Naoi, Taketo, David Airey, Shoji Iijima, and Outi Niininen. 2006. "Visitors' Evaluation of an Historical District: Repertory Grid Analysis and Laddering Analysis with Photographs." *Tourism Management* 27, 420–36.

Newman, Bruce I., and Jagdish N. Sheth. 1985. "A Model of Primary Voter Behavior." *Journal of Consumer Research* 12 (September), 178–87.

Nielsen, Niels A., Tino Bech-Larsen, and Klaus G. Grunert. 1998. "Consumer Purchase Motives and Product Perceptions: A Laddering Study on Vegetable Oil in Three Countries." *Food Quality and Preference* 9 (6), 455–66.

Nisbett, Richard E., and Timothy DeCamp Wilson. 1977. "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84 (May), 234–59.

Norton, John. 1987. "Reagan-Bush '84." Case No. UVA-M-340. Charlottesville: University of Virginia Darden School Foundation.

Olson, Jerry C., and Thomas J. Reynolds. 2001. "The Means-End Approach to Understanding Consumer Decision Making." In *Understanding Consumer Decision Making: The Means-End Approach to Marketing and Advertising Strategy*, ed. Thomas J. Reynolds and Jerry C. Olson. Mahwah, NJ: Lawrence Erlbaum.

Orsingher, Chiara, and Gian Luca Marzocchi. 2003. "Hierarchical Representation of Satisfactory Consumer Service Experience." *International Journal of Service Industry Management* 14 (2), 200–216.

Overby, Jeffrey W., Sarah F. Gardial, and Robert B. Woodruff. 2004. "French Versus American Consumers' Attachment of Value to a Product in a Common Consumption Context: A Cross-National Comparison." *Journal of the Academy of Marketing Science* 32 (4), 437–60.

Parry, Mark E. 2002. *Strategic Marketing Management: A Means-End Approach*. New York: McGraw-Hill Executive MBA Series.

Peffers, Ken, Charles E. Gengler, and Tuure Tuunanen. 2003. "Extending Critical Success Factors Methodology to Facilitate Broadly Participative Information Systems Planning." *Journal of Management Information Systems* 20 (1), 51–85.

Perreault, Jr., William D. and Laurence E. Leigh. 1989. "Reliability of Nominal Data Based on Qualitative Judgments." *Journal of Marketing Research* 26 (May), 135–48.

Pieters, Rik, Hans Baumgartner, and Doug Allen. 1995. "A Means-End Approach to Consumer Goal Structures." *International Journal of Research in Marketing* 12 (October), 227–44.

Pieters, Rik, Günther Bottschen, and Eva Thelen. 1998. "Customer Desire Expectations About Service Employees: An Analysis of Hierarchical Relations." *Psychology and Marketing* 15 (8), 755–73.

Phillips, Joan M. and Thomas J. Reynolds. 2009. "A Hard Look at Hard Laddering: A Comparison of Studies Examining the Hierarchical Structure of Means-End Theory." *Qualitative Market Research: An International Journal* 12, forthcoming.

Phillips, Joan M., Joel E. Urbany, and Thomas J. Reynolds. 2008. "Confirmation and the Effects of Valenced Political Advertising: A Field Experiment." *Journal of Consumer Research* 34 (April), 794-806.

Quattrone, George A. 1985. "On the Congruity Between Internal States and Action." *Psychological Bulletin* 98 (1), 3–40.

Reynolds, Thomas J. 2006. "Methodological and Strategic Development Implications of Decision Segmentation." *Journal of Advertising Research* 46 (December), 445-61.

Reynolds, Thomas J., and Alyce B. Craddock. 1988. "The Application of the MECCAS Model to the Development and Assessment of Advertising Strategy: A Case Study." *Journal of Advertising Research* 28 (April/May), 43–54.

Reynolds, Thomas J., Clay Dethloff, and Steven J. Westberg. 2001. "Advancements in Laddering." In *Understanding Consumer Decision Making: The Means-End Approach to Marketing and Advertising Strategy*, ed. Thomas J. Reynolds and Jerry C. Olson. Mahwah, NJ: Lawrence Erlbaum.

Reynolds, Thomas J., and Jonathan Gutman. 1988. "Laddering Theory, Method, Analysis, and Interpretation." *Journal of Advertising Research* 28 (February/March), 11–31.

Reynolds, Thomas J., and Jerry C. Olson, eds. 2001. *Understanding Consumer Decision Making: The Means-End Approach to Marketing and Advertising Strategy.* Mahwah, NJ: Lawrence Erlbaum.

Reynolds, Thomas J., and Minakshi Trivedi. 1989. "An Investigation of the Relationship Between the MECCAS Model and Advertising Affect." In *Advertising and Consumer Psychology*, ed. Alice Tybout and Pat Cafferata. Vol. 4. Lexington, MA: Lexington Books.

Reynolds, Thomas J., and David B. Whitlark. 1995. "Applying Laddering Data to Communications Strategy and Advertising Practice." *Journal of Advertising Research* 35 (4), 9–17.

Reynolds, Thomas J., David B. Whitlark, and Richard B. Wirthlin. 2001. "Effectively Translating In-Depth Consumer Understanding into Communications Strategy and Advertising Practice." In *Understanding Consumer Decision Making: The Means-End Approach to Marketing and Advertising Strategy*, ed. Thomas J. Reynolds and Jerry C. Olson. Mahwah, NJ: Lawrence Erlbaum.

Roininen, Katariina, Anne Arvola, and Liisa Lähteenmäki. 2006. "Exploring Consumers' Perceptions of Local Food with Two Different Qualitative Techniques: Laddering and Word Association." *Food Quality and Preference* 17, 20–30.

Roininen, Katariina, Laurence Fillion, David Kilcast, and Liisa Lähteenmäki. 2004. "Exploring Difficult

Textural Properties of Fruit and Vegetables for the Elderly in Finland and the United Kingdom." *Food Quality and Preference,* 15 (6), 517–530.

Roininen, Kateriina, Liisa Lähteenmäki, and Hely Tuorila. 2000. "An Application of Means-End Chain Approach to Consumers' Orientation to Health and Hedonic Characteristics of Foods." *Ecology of Food and Nutrition* 39 (1), 61–81.

Rokeach, Milton J. 1973. *The Nature of Human Values.* New York: The Free Press.

Roth, Martin S. 1994. "Enhancing Consumer Involvement in Health-Care—The Dynamics of Control, Empowerment, and Trust." *Journal of Public Policy and Marketing* 13 (1), 115–32.

Rothschild, Michael L. 1978. "Political Advertising: A Neglected Policy Issue in Marketing." *Journal of Marketing Research* 15 (February), 58–71.

Russell, C.G., A. Busson, I. Flight, J. Bryan, J.A. van Lawick van Pabst, D.N. Cox. 2004a. "A Comparison of Three Laddering Techniques Applied to an Example of a Complex Food Choice." *Food Quality and Preference* 15, 569–83.

Russell, C.G., I. Flight, P. Leppard, J.A. van Lawick van Pabst, J.A. Syrette, and D.N. Cox. 2004b. "A Comparison of Paper-and-Pencil and Computerised Methods of 'Hard' Laddering." *Food Quality and Preference* 15, 279–91.

Rust, Roland T. 2004. "If Everything Is Service, Why Is This Happening Now, and What Difference Does It Make?" *Journal of Marketing* 68 (January), 23–24.

Rust, Roland T., and Bruce Cooil. 1994. "Reliability Measures for Qualitative Data: Theory and Implications." *Journal of Marketing Research* 31 (February), 1–14.

Schuman, Howard, and Stanley Presser. 1996. *Questions and Answers in Attitude Surveys.* Thousand Oaks, CA: Sage.

Shugan, Steven M. 2004. "Finance, Operation, and Marketing Conflicts in Service Firms." *Journal of Marketing* 68 (January), 24–26.

Simmons, Carolyn J., Barbara A. Bickart, and John G. Lynch, Jr. 1993. "Capturing and Creating Public Opinion in Survey Research." *Journal of Consumer Research* 20 (September), 316–29.

Sirsi, Ajay K., James C. Ward, and Peter H. Reingen. 1996. "Microcultural Analysis of Variation in Sharing of Causal Reasoning About Behavior." *Journal of Consumer Research* 22 (March), 345–72.

Snelders, Dirk, and Jan P.L. Schoormans. 2004. "An Exploratory Study of the Relation Between Concrete and Abstract Product Attributes." *Journal of Economic Psychology* 25, 803–20.

Søndergaard, Helle Alsted. 2005. "Market-Oriented New Product Development: How Can a Means-End Chain Approach Affect the Process?" *European Journal of Innovation Management* 8 (1), 79–90.

Subramony, Deepak Prem. 2002. "Why Users Choose Particular Web Sites over Others: Introducing a 'Means-End' Approach to Human-Computer Interaction." *Journal of Electronic Commerce Research* 3 (3), 144–61.

Tan, Felix, and R. Brent Gallupe. 2006. "Aligning Business and Information Systems Thinking: A Cognitive Approach." *IEEE Transactions on Engineering Management* 53 (May) 223–37.

Taylor, Stephanie D., Richard P. Bagozzi, Caroline A. Gaither, and Kenneth A. Jamerson. 2006. "The Bases of Goal Setting in the Self-Regulation of Hypertension." *Journal of Health Psychology* 11 (1), 141–62.

ter Hofstede, Frenkel, Anke Audenaert, Jan-Benedict E. M. Steenkamp, and Michel Wedel. 1998. "An Investigation into the Association Pattern Technique as a Qualitative Approach to Measuring Means-End Chains." *International Journal of Research in Marketing* 15 (1), 37–50.

ter Hofstede, Frenkel, Jan-Benedict E.M. Steenkamp, and Michel Wedel. 1999. "International Market Segmentation Based on Consumer-Product Relations." *Journal of Marketing Research* 36 (February), 1–17.

Thompson, Craig J., and Maura Troester. 2002. "Consumer Value Systems in the Age of Postmodern Fragmentation: The Case of the Natural Health Microculture." *Journal of Consumer Research* 28 (March), 550–71.

Urala, Nina, and Liisa Lähteenmäki. 2003. "Reasons Behind Consumers' Functional Food Choices." *Nutrition and Food Science* 33 (4), 148–58.

Valette-Florence, Pierre. 1998. "A Causal Analysis of Means-End Hierarchies in a Cross-Cultural Context: Methodological Refinements." *Journal of Business Research* 42, 161–66.

Valette-Florence, Pierre, and Bernard Rapacchi. 1991. "Improvements in Means-End Chain Analysis: Using Graph Theory and Correspondence Analysis." *Journal of Advertising Research* (February/March), 30–45.

Vannoppen, Jan, Wim Verbeke, and Guido Van Huylenbroeck. 2002. "Consumer Value Structures Towards Supermarket Versus Farm Shop Purchase of Apples from Integrated Production in Belgium." *British Food Journal* 104 (10), 828–44.

van Rekom, Johan, Cees B.M. van Riel, and Berend Wierenga. 2006. “A Methodology for Assessing Organizational Core Values.” *Journal of Management Studies* 43 (2), 175–201.
Vargo, Stephen L., and Robert F. Lusch. 2004. “Evolving to a New Dominant Logic for Marketing.” *Journal of Marketing* 68 (January), 1–17.
Veryzer, Robert W. 1998. “Key Factors Affecting Customer Evaluation of Discontinuous New Products.” *Journal of Product Innovation Management* 15 (2), 136–50.
———. 1999. “A Nonconscious Processing Explanation of Consumer Response to Product Design.” *Psychology and Marketing* 16 (6), 497–522.
Walker, Beth A., and Jerry C. Olson. 1991. “Means-End Chains: Connecting Products with Self.” *Journal of Business Research* 22 (March), 111–18.
Wansink, Brian. 2003. “Using Laddering to Understand and Leverage a Brand’s Equity.” *Qualitative Market Research: An International Journal* 6 (2), 111–18.
Wendel, Sonja, and Benedict G.C. Dallaert. 2005. “Situation Variation in Consumers’ Media Channel Consideration.” *Journal of the Academy of Marketing Science* 33 (Fall), 575–84.
White, Helen, and Katerini Kokotsaki. 2004. “Indian Food in the UK: Personal Values and Changing Patterns of Consumption.” *International Journal of Consumer Studies* 28 (3), 284–94.
Wirthlin, Dick. 2004. *The Greatest Communicator: What Ronald Reagan Taught Me About Politics, Leadership, and Life.* Hoboken, NJ: John Wiley & Sons.
Woodruff, Robert B. 1997. “Customer Value: The Next Source for Competitive Advantage.” *Journal of the Academy of Marketing Science* 25 (2), 139-53.
Zaltman, Gerald, and Robin Higie Coulter. 1995. “Seeing the Voice of the Customer—Metaphor-based Advertising Research.” *Journal of Advertising Research* 35 (4), 35–51.
Zanoli, Raffaele, and Simona Naspetti. 2002. “Consumer Motivations in the Purchase of Organic Food: A Means-End Approach.” *British Food Journal* 104 (8), 643–53.
Zeithaml, Valerie. 1988. “Consumer Perceptions of Price, Quality, and Value: A Means-End Model of Synthesis of Evidence.” *Journal of Marketing* 52 (July), 2–22.
Zwarts, Astrid, and Henny Coolen. 2006. “The Meaning of Preferences for Residential Environment Features: A Case Study Among Apartment Dwellers in the Netherlands.” *Journal of Architectural and Planning Research* 23 (3), 200–215.

Appendix A. Glossary of Laddering Terms

Attribute—a defining characteristic of an object that derives its importance from one’s values.

Code-pair—refers to two adjacent elements. The “connected significance” of a code-pair is determined by the threshold cutoff level.

Direct connections—refers to implicative relations among adjacent elements. Note that the direct implications AB, BC, CD, XB, BY, and YZ also “imply” these direct and indirect implications: AC, AD, BD, XY, XZ, and BZ as well as AY, AZ, BZ, XC, XD, and BD.

Element—a code within a means-end chain that represents a collection of similar or identical elicitations found within a group of ladders.

Functional Consequence—a salient, rather tangible, and direct experience that is likely to occur immediately after a decision. May be positive or negative. Positive functional consequences are “benefits.” Roughly half are negative, which translates into the basis of a decision being avoidance driven.

Hierarchical value map (HVM)—a tree diagram (directed graph or dendogram) consisting of a series of means-end chains constructed from the significant code-pair connections.

Implications—the number of connections, or linkages, between elements.

Indirect connections—refers to implicative relations among nonadjacent elements.

Ladder—the sequence of elicitations obtained from individual respondents in a laddering interview. A "complete" ladder consists of an elicitation at each level of abstraction. Many researchers use the terms ladder and MEC interchangeably. The following is an illustration of a complete ladder/MEC with five elements (codes).

Value→	**Security self/children's future**
Implication→	↑
Psychosocial consequence→	**Cares and concerned about people**
Implication→	↑
Psychosocial consequence→	**Feel Sympathetic**
Implication→	↑
Functional consequence→	**Fairness**
Implication→	↑
Attribute→	**Reduce poverty**

Laddering—an in-depth, one-on-one interviewing technique used to develop an understanding of how consumers translate the attributes of product, services, etc. into meaningful association with respect to self, following means-end theory.

Levels of abstraction—a series of four sequentially related levels of (higher-order) meanings (i.e., attribute→functional consequence→psychosocial consequence→value).

Means-end chain (MEC)—the sequence of individual elements that emerges from laddering representing the levels of abstraction or may refer to an aggregation of individual ladders. Many researchers use the terms MEC and ladder interchangeably.

Means-end theory—premised on the belief that decisions are based on one's desired outcome or end-state, and focuses on the linkages between the attributes that exist (the "means"), the consequences provided by the attribute, and the personal values (the "ends") these consequences reinforce (Gutman 1982).

Psychosocial consequence—the personal, sometimes emotional, outcome that the functional consequence provides.

Threshold cutoff level—determines the "significant," or dominant, code-pairs to be included in the hierarchical value map. This determination typically captures 70 percent of the implications in the data to avoid the infrequent, less meaningful connections. Therefore, a threshold cutoff of 3 would include only those code-pairs that occur in 3 or more individual ladders.

Value—a motivating "end-state of existence" that individuals strive for in their lives. Oftentimes referred to in a choice context as a "driving force."

Appendix B. Computation of Laddering Quality Metrics

To illustrate the laddering methodology incorporating the suggested metrics, the following score matrix for data set A consisting of twelve complete ladders is presented below. The numerical designations within levels (a1–a4 = attributes, f1–f5 = functional consequences, p1–p3 = psychosocial consequences, and v1–v2 = values) represent individual codes that total 14 overall.

Data Set A: Score Matrix

Ladder	Element 1	Element 2	Element 3	Element 4	Element 5	Length
1	a1	f1	f5	p2	v2	5
2	a1	f1	p2	v2		4
3	a1	f2	p2	v2		4
4	a1	f3	p2	v1		4
5	a2	f2	f5	p2	v2	5
6	a2	f4	p2	v2		4
7	a2	f2	p1	v1		4
8	a3	f2	f5	p2	v2	5
9	a3	f3	p3	v2		4
10	a3	f4	p2	v2		4
11	a4	f2	f5	p2	v1	5
12	a4	f5	p1	v2		4

Note that average ladder length = 4.3 and LQ = 100 percent (12 complete ladders/12 total ladders).

An implication matrix is generated from the score matrix. The numbers to the left side of the decimal are direct connections (a directly preceded b in the data). The numbers to the right side of the decimal (hundredths) summarize the number of indirect connections in the data (a preceded b in the ladders, but not directly). This presentation heuristic of separating direct and indirect implication by the decimal is for illustrative purposes. (Note: there are no implications within attributes (a's) or within values (v's), so these columns have been excluded.)

Data Set A: Implication Matrix

	f1	f2	f3	f4	f5	p1	p2	p3	v1	v2	Σ	**Σ sig**
a1	**2.00**	1.00	1.00	0.01			**0.04**		0.01	**0.03**	13	**9**
a2		**2.00**		**1.01**		0.01	**0.02**		0.01	**0.02**	10	**8**
a3		1.00	1.00	**1.01**			**0.02**	0.01		**0.03**	10	**7**
a4		1.00		0.01	1.00	0.01	0.01		0.01	0.01	7	**0**
f1				1.00			**1.01**			**0.02**	5	**4**
f2				**3.00**		1.00	**1.03**		**0.02**	**0.03**	13	**12**
f3							1.00	1.00	0.01	0.01	4	**0**
f4							**6.00**		0.01	**0.05**	12	**11**
f5						1.00				0.01	2	**0**
p1									1.00	1.00	2	**0**
p2									**2.00**	**7.00**	9	**9**
p3										1.00	1	**0**
Σ	2	5	2	10	1	4	22	2	10	30	88	**60**

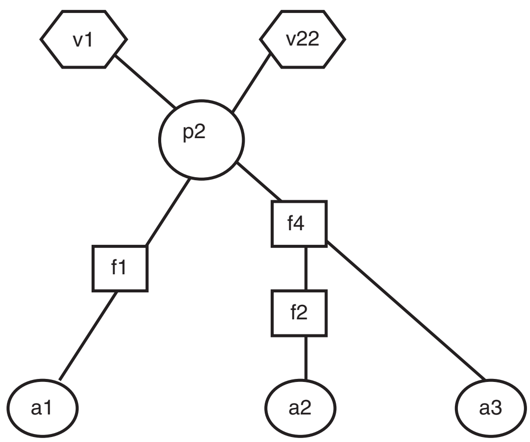
The decision as to which cutoff level is most appropriate to summarize the data results from an analysis of how many total implications are accounted for at the varying levels.

Data Set A: Proportion of "Significant" Implications by Cutoff Level

Cutoff Level	*Quasi*-Coefficient of Determination r^2_q	←Systematic variance
1	1.000	←All implications are significant at a cutoff of one
2	**.682**	←Cutoff selected; 68.2% of implications are accounted for
3	.432	

The decision to select a cutoff = 2 accounting for approximately 70 percent of the total connection is consistent with traditional laddering research methodology. The total connections greater than or equal to 2 are **bold** in the implication matrix above. A directed graph (HVM) is constructed that embodies all of the significant connections in as parsimonious manner as possible (Reynolds and Gutman 1988). The construction of the map in this manner necessarily embeds the significant connections in as simple a graphical structure as possible to aid in interpretation.

Summary HVM for Example Laddering Data A
(12 ladders; cutoff = 2; 68.2 percent of implications)



The computation of the *quasi*-reliability metric, r_q, for 68.2 percent (60/88) of the connections represented in this HVM (at the cutoff level of 2) is the square root of the *quasi*-coefficient of determination, $r^2{}_q$, considered the proportion of systematic variance accounted for, or .826.

Contrasting of Two Laddering Solutions: Quasi-Validity (v_q)

The question of comparing two means-end solutions (and their resulting HVMs) involves the computation of the *quasi*-validity measure, v_q, which reflects the common variance between two solutions at their respective cutoff levels. For example, consider the example data and summary map for data set A above. The research question is: To what degree (percentage of total implications of A or 88) are the significant pairwise connections in the solution for A consistent with the significant pairwise connections for a given solution of data set B (presented below)? Note that average ladder length = 4.4 for this data set and that and L_Q = 100 percent.

Data Set B: Score Matrix

Ladder	Element 1	Element 2	Element 3	Element 4	Element 5	Length
1	a1	f1	f4	p2	v2	5
2	a1	f1	f4	v1		4
3	a1	f2	p2	v2		4
4	a1	f3	p2	v1		4
5	a2	f2	f4	p2	v2	5
6	a2	f4	p2	v2		4
7	a2	f2	p1	v2		4
8	a3	f2	f4	p2	v2	5
9	a3	f3	p3	v2		4
10	a3	f4	p2	v2		4
11	a4	f2	f4	p2	v1	5
12	a4	f5	p1	v1		4
13	a1	f1	f4	p2	v1	5
14	a1	f1	f4	p3	v2	5
15	a1	f4	p2	v1		4
16	a2	f2	f4	p2	v1	5
17	a2	f4	p2	v1		4
18	a2	f2	p2	v2		4
19	a3	f2	p2	v2		4
20	a3	f2	f4	p1	v2	5
21	a3	f2	f4	p1	v2	5
22	a4	f3	f5	v2		4
23	a4	f5	p2	v2		4
24	a4	f1	p3	v2		4

Using a decision heuristic to "cutoff" at or below a level representing 70 percent of total implications, we select a cutoff level of 4.

Data Set B: Proportion of "Significant" Implications by Cutoff Level

Cutoff Level	*Quasi*-Coefficient of Determination r^2_q	
		←Systematic variance
1	1.000	←All implications are significant at a cutoff of one
2	.883	
3	.761	
4	**.628**	←Cutoff selected; 62.8% of implications are accounted for

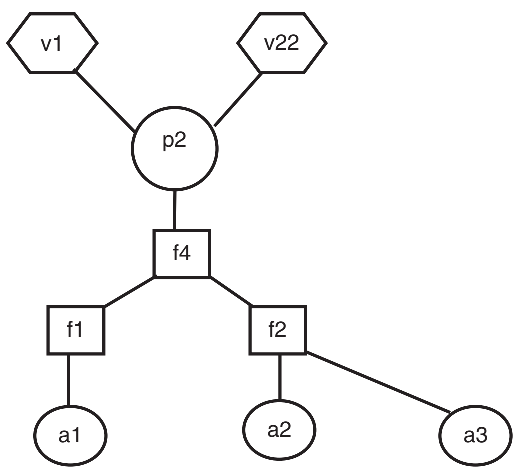
The "significant" connections for data set B are **bold** in the implication matrix below.

Data Set B: Implication Matrix

	f1	f2	f3	f4	f5	p1	p2	p3	v1	v2	Σ	**ΣSig**
a1	**4.00**	1.00	1.00	**1.04**			0.05	0.01	**0.04**	0.03	24	**18**
a2		4.00		2.02		0.01	**0.05**		0.02	**0.04**	20	**17**
a3		**4.00**	1.00	**1.03**		0.02	0.03	0.01		**0.06**	21	**14**
a4	1.00	1.00	1.00	0.01	2.01	0.01	0.02	0.01	0.02	0.03	16	**0**
f1				4.00			0.02	1.01	0.02	0.03	13	**4**
f2				6.00		1.02	**3.04**		0.02	**0.08**	26	**21**
f3					1.00		1.00	1.00	0.01	0.02	6	**0**
f4						2.00	**10.0**	1.00	**1.05**	**0.08**	27	**24**
f5						1.00	1.00		0.01	1.01	5	**0**
p1									1.00	3.00	4	**0**
p2									**6.00**	**9.00**	15	**15**
p3										3.00	3	**0**
Σ	5	10	3	24	4	10	36	7	27	54	180	**113**

The HVM for Data Set B is presented below.

Summary HVM for Example Laddering Data B
(24 ladders; cutoff = 4; 62.8% of implications)



The computation of the *quasi*-reliability metric, r_q, for 62.8 percent of the connections (113/180) represented in this HVM (at the cutoff level of 4) is the square root of what is considered the systematic variance accounted for, or .792.

A visual comparison of the two resulting HVMs appears to have one relatively small difference and one fairly large one. For data set B, the connection of a3 linking to f2 rather than to f4 directly (as in A) represents a small difference between the two solutions. The larger difference in B, as contrasted to A, is that neither a1 nor f1 is connected to f4.

To compute the *quasi*-validity measure, v_q, reflecting the goodness-of-fit of A with respect to B, the percentage of the significant common implications of A with B relative to the total implications of A is computed. This is accomplished by first summarizing the significant connections in both A and B (A, B) at their respective cutoff levels in the summary table below. Recall that the total implications for each data set (from the implication matrices above) are: A = 88 (12 ladders) and B = 180 (24 ladders), respectively.

Summary Table of Common Significant Implications

	f1	f2	f3	f4	f5	p1	p2	p3	v1	v2	Σ **sigA**	Σ **sigB**
a1	**2, 4**			0, 5			**4, 5**		0, 4	3, 0	6	9
a2		**2, 4**		**2, 4**			**2, 5**			**2, 4**	8	17
a3		0, 4		**2, 4**			2, 0			**3, 6**	5	10
a4											0	0
f1				0, 4			2, 0			2, 0	0	0
f2				**3, 6**			**4, 7**		2, 0	**3, 8**	10	21
f3											0	0
f4							**6, 10**		0, 6	**5, 8**	11	18
f5											0	0
p1											0	0
p2									**2, 6**	**7, 9**	9	15
p3											0	0
Σ											49	90

Cells for connections that are not significant in either data set are empty. For example, the link from a4 to f1 is not significant in the implication matrices for A or B. Cells with a non-bolded entry indicate that the implication is significant in one data set, but not the other. For example, the connection from a3 to f2 is significant in data set B, but not in A. **Bold** entries are the "common" significant implications. That is, these entries significant for both A and B.

The computation of the *quasi*-validity measure, v_q, of A relative to B is then obtained by totaling the common significant connections (i.e., bolded and both nonzero) for A (49) and dividing by the total number of implications in A's implication matrix (88), or, .556. (For the sake of completeness, the *quasi*-validity measure, v_q, for B relative to A is the total of the bolded, nonzero implications for B (90) divided by the total number of implications in B's implication matrix (180), or, .500. Given the size of the respective data sets (A = 12 ladders, B = 24 ladders), it would be most relevant to assess the "fit" by contrasting the smaller data set (A) to the larger data set (B).

As this illustration demonstrates, simply contrasting the similarity of the resulting HVMs does not provide a clear basis to quantify the "fit" of the respective data sets given that the primary goal of HVM construction is to create as simple a structure as possible. Only by knowing the significant common connections in both data sets, which are embedded in the HVM relative to the total number of implications, can a measure of "fit," termed *quasi*-validity, be calculated.

The following table is a summary of the quality metrics calculated in the exercise above.

Summary of Laddering Quality Metrics

	Ladder Quality	*Quasi*-Coefficient of Determination	*Quasi*-Reliability	*Quasi*-Validity
	L_Q	$r^2{}_q$	r_q	r_q
Data set A	100%	.682	.826	.556
Data set B	100%	.628	.792	.500

Appendix C. Code Distributions

	Bush Top of Mind		Kerry Top of Mind		Bush Disequity		Kerry Disequity	
	F-to-F	Online	F-to-F	Online	F-to-F	Online	F-to-F	Online
Attributes								
Average citizen orientation	.04	.08	.02	.05	.00	.10	.03	.07
Change in office	.01	.00	.09	.10	.00	.00	.03	.00
Aggressive foreign policy	.39	.30	.04	.00	.39	.33	.06	.08
Lack of clear position	.01	.01	.24	.22	.05	.00	.46	.29
Intelligence	.13	.15	.05	.04	.07	.03	.03	.02
Military record	.00	.00	.11	.03	.00	.00	.02	.08
Liberal Democrat	.00	.00	.13	.14	.02	.00	.06	.14
Individual rights	.01	.00	.01	.04	.00	.00	.17	.10
Society rights	.03	.05	.00	.00	.08	.03	.00	.02
Candidate image	.35	.26	.30	.38	.17	.22	.13	.19
Conservative Republican	.03	.15	.00	.00	.22	.28	.00	.02
Level totals	72	74	82	78	59	60	63	59
	$\chi^2 = 11.61$, df = 8, n.s.		$\chi^2 = 9.93$, df = 8, n.s.		$\chi^2 = 13.08$, df = 7, $p < .10$		$\chi^2 = 14.07$, df = 10, n.s.	
Functional Consequences								
U.S. issues need attention	.03	.06	.07	.05	.12	.13	.06	.05
Leadership ability	.27	.25	.28	.29	.17	.17	.22	.19
Experience	.03	.02	.02	.05	.00	.00	.00	.00
U.S. military lives	.07	.10	.05	.01	.15	.11	.02	.03
Trustworthy	.13	.21	.22	.15	.07	.19	.25	.27
Voter representation	.04	.08	.06	.08	.07	.11	.03	.03
U.S. economy	.13	.03	.14	.13	.15	.06	.14	.06
Principles values	.09	.08	.08	.07	.11	.06	.19	.19
U.S. world standing	.10	.08	.02	.07	.09	.01	.08	.08
Caring	.10	.08	.06	.10	.08	.16	.02	.11
Level totals	90	87	86	86	75	70	64	64
	$\chi^2 = 9.32$, df = 9, n.s.		$\chi^2 = 7.61$, df = 9, n.s.		$\chi^2 = 15.18$, df = 8, $p < .10$		$\chi^2 = 7.08$, df = 8, n.s.	

Psychosocial Consequences								
In control	.27	.14	.22	.09	.34	.18	.36	.11
Defense orientation	.08	.05	.09	.01	.08	.06	.04	.09
Informed need	.01	.03	.07	.01	.01	.05	.12	.06
Personal freedom	.14	.05	.08	.07	.12	.09	.09	.08
Confidence	.26	.39	.31	.57	.23	.42	.17	.38
Economic opportunity	.10	.05	.08	.13	.10	.03	.04	.06
American unity	.06	.13	.03	.04	.03	.05	.07	.09
U.S. society well-being	.08	.16	.13	.08	.09	.12	.10	.13
Level totals	119	77	105	76	105	65	98	64
	$\chi^2 = 18.01$, df = 7, $p < .05$		$\chi^2 = 20.75$, df = 7, $p < .01$		$\chi^2 = 17.79$, df = 7, $p < .05$		$\chi^2 = 19.42$, df = 7, $p < .01$	
Values								
Quality of life	.03	.19	.04	.25	.06	.17	.07	.13
Peace of mind	.27	.25	.40	.42	.29	.32	.36	.31
Personal security	.32	.28	.37	.21	.35	.27	.29	.27
Independence	.07	.08	.00	.03	.08	.05	.02	.03
Patriotism	.10	.13	.04	.04	.02	.07	.07	.13
Family security	.07	.04	.03	.03	.04	.07	.07	.08
Self-esteem	.15	.00	.10	.04	.15	.00	.13	.03
Belonging	.00	.03	.01	.00	.02	.05	.00	.02
Level totals	60	75	70	77	52	59	56	62
	$\chi^2 = 20.49$, df = 7, $p < .01$		$\chi^2 = 18.59$, df = 7, $p < .01$		$\chi^2 = 15.59$, df = 7, $p < .05$		$\chi^2 = 6.66$, df = 7, n.s.	
Grand totals	341	313	343	317	291	254	281	249

Appendix D. Summary of Quasi-Reliability and Quasi-Validity Measures by Laddering Method

					Threshold Cutoff = 3				Threshold Cutoff = 4				Threshold Cutoff = 5			
					Quasi- Reliability		*Quasi-* Validity		*Quasi-* Reliability		*Quasi-* Validity		*Quasi-* Reliability		*Quasi-* Validity	
	# Ladders		# Implications		(r_q)		(v_q)		(r_q)		(v_q)		(r_q)		(v_q)	
Ladder type	F	O	F	O	F	O	F	O	F	O	F	O	F	O	F	O
Bush TOM	72	72	660	535												
All (direct + indirect)					.87	.81	.52	.70	.84	.65	.42	.90	.78	.60	.41	.82
CoD_q (r_q^2)					.77	.66	—	—	.71	.42	—	—	.61	.36	—	—
Direct only					.82	.76	.53	.65	.76	.63	.46	.78	.69	.57	.44	.72
CoD_q (r_q^2)					.67	.59	—	—	.58	.40	—	—	.48	.33	—	—
Kerry TOM	72	72	671	554												
All (direct + indirect)					.87	.83	.59	.90	.80	.77	.54	.75	.68	.68	.54	.74
CoD_q (r_q^2)					.76	.69	—	—	.64	.59	—	—	.46	.46	—	—
Direct only					.82	.80	.59	.77	.74	.74	.54	.73	.66	.65	.50	.77
CoD_q (r_q^2					.68	.64	—	—	.55	.55	—	—	.43	.42	—	—
Bush Disequity	62	59	558	431												
All (direct + indirect)					.88	.81	.47	.75	.77	.70	.36	.66	.69	.64	.42	.67
CoD_q (r_q^2)					.77	.66	—	—	.59	.49	—	—	.48	.41	—	—
Direct only					.79	.77	.44	.56	.69	.66	.35	.55	.61	.60	.38	.50
CoD_q (r_q^2)					.63	.60	—	—	.48	.43	—	—	.37	.36	—	—
Kerry Disequity	61	58	521	417												
All (direct + indirect)					.83	.76	.55	.70	.77	.63	.40	.66	.68	.53	.38	.70
CoD_q (r_q^2)					.69	.56	—	—	.59	.40	—	—	.46	.28	—	—
Direct only					.76	.71	.47	.54	.69	.60	.41	.54	.60	.49	.33	.47
CoD_q (r_q^2)					.57	.50	—	—	.47	.36	—	—	.37	.24	—	—

Note: # Ladders = number of ladders; # Implications = total number of implications; F = face-to-face; O = online methods; "direct only" connections include those significant code-pairs meeting or exceeding the threshold value; direct + indirect includes those code-pairs that are also implied by the significant direct code-pairs; CoDq (rq2) = quasi-coefficient of determination is the proportion of implications classified as systematic variance.