Description of a Northern California Shopping Survey

Data Collection Effort

David T. Ory
Department of Civil and Environmental Engineering
and
Institute of Transportation Studies
University of California, Davis
Davis, CA 95616
voice: (415) 378-9102
fax: (530) 752-6572
e-mail: dtory@ucdavis.edu

and

Patricia L. Mokhtarian
Department of Civil and Environmental Engineering
and
Institute of Transportation Studies
University of California, Davis
Davis, CA 95616
voice: (530) 752-7062
fax: (530) 752-7872
e-mail: plmokhtarian@ucdavis.edu

March 2007

This research is funded by the University of California Transportation Center.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Survey design and data collection was funded by the University of California Transportation Center. Early phases of the survey design were conducted by Xinyu Cao; we greatly appreciate his thoughtful and extensive input to the final design. Susan Handy and numerous pretesters also offered many useful suggestions for improvement. Tara Puzin provided able assistance in the identification and analysis of prospective survey census tracts.

1. INTRODUCTION

Applications of new information and communication technologies (ICTs) are changing how and where we work, shop, play, travel, and in other ways live our lives. Yet because ICT development and use is in such a volatile state, many of those changes and impacts are poorly understood. This report summarizes the development and deployment of a survey instrument intended to gather information to allow better understanding of the transportation impacts of business-to-consumer (B2C) e-commerce. Although the business-to-business (B2B) segment dominates e-commerce in terms of the dollar value of transactions made, B2C remains important for its potential impacts on urban travel and land use patterns, including potential redistributions of retail land uses, and substantial increases of package delivery trips into residential neighborhoods.

We see an analysis of the transportation impacts of B2C e-commerce as having two components: (1) assessing the transportation impacts of a given level or pattern of adoption of B2C e-commerce, and (2) investigating the adoption of B2C e-commerce (who, under what circumstances, in what form). While transportation planners ultimately need to forecast transportation impacts, to do so accurately they need to understand adoption processes and trends. Thus, the data collection described here is intended to lead to modeling the adoption of B2C e-commerce, among other shopping "modes" (specifically store and catalog shopping). Because we take the consumer perspective, we will refer to the use of the internet for B2C e-commerce as "e-shopping". We consider e-shopping to be a subset of "teleshopping", which also includes catalog shopping (whether placing the order by phone or mail) and shopping from a television channel (generally by phone). To frame a manageable project, and because most TV shopping appears to be impulsive (Handy and Yantis, 1997), we do not include TV shopping in the survey.

Though the survey instrument collects data on the pre-purchase browsing mode(s) as well as the transaction mode choice, our definition of shopping requires a purchase to occur, not just information-gathering or "window shopping". Because (1) the nature of the shopping process is presumed to be quite heterogeneous depending on the type of good being considered; (2) due to resource constraints our sample will number in the (high) hundreds rather than many thousands; and (3) attempting to survey respondents in-depth with respect to many types of goods would pose too great a burden; we focus on two specific product types. Product classes can be characterized along a number of dimensions, including frequency of purchase, cost, search area (local, regional, broader), tangibility (whether a service, a digital good, or a material good), perishability, "differentiability" (the extent to which retailers can distinguish their offering of the product from others'), product information complexity, and as a "search good" (having features "that can be evaluated from externally provided information") versus an "experience good" (needing "to be personally inspected or tried"; Peterson, et al., 1997).

To yield a sufficient number of purchase occasions in our sample, we focus on low-cost, medium-/high-frequency-of-purchase goods. Specifically, we take the tangible product classes of books/physical CDs/DVDs/videotapes and clothing/shoes to be analyzed in depth. Books/CDs/DVDs/videotapes per se are basically search goods with low intrinsic differentiability, though the bookstore experience (browsing contents, having coffee) may offer some advantages to some customers. Clothing and shoes are basically experience goods, although decades of catalog shopping history show that some customers are willing to buy without trying. The survey also asks a few questions about internet shopping activity of all kinds, especially internet-based purchases.

The main research questions to be addressed by this exploratory project build on a number of prior studies: (1) For the product class in question, what are the advantages (motivators and facilitators for choosing) and disadvantages of (constraints on choosing) each shopping mode? A number of authors have identified the potential advantages of e-shopping and/or store shopping (e.g. Brynjolfsson and Smith, 2001; Salomon and Koppelman, 1988; Tauber, 1972). Store shopping is so far still very different from e-

shopping in terms of such attributes as the information provided, the sensual stimulation, the ability to compare prices and to attain immediate ownership. Beyond the functional attributes related to shopping and purchasing, store shopping also offers numerous other experiences, which to varying degrees are less amenable to electronic platforms. These include, for example, the ability to interact with real salespeople and to bargain, the opportunity to be outside the home or work environment, some physical exercise, and so on. Shopping, under many circumstances, is a combined maintenance - leisure activity. Thus, the choice between store shopping and e-shopping is not unambiguous (Handy and Yantis, 1997).

(2) Can market segments with different propensities to use alternative modes be identified? A composite of several studies (Cairns, 1996; Koppelman, et al., 1991; Tacken, 1990; Gould and Golob, 1997; Gould et al., 1998; Burke, 1998; Farag et al., 2006; Ferrell, 2005; Ren and Kwan, 2005) identifies four segments of the population that are likely to be early adopters of e-shopping: the mobility-limited, time-starved, technophilic, and shopping-phobic. These four segments may well be generalizable to many e-shopping contexts. In reality, however, they probably represent (and in general we will treat them as) four continuous dimensions, with individuals separately falling somewhere along each of them, rather than four mutually-exclusive and collectively-exhaustive group membership indicators.

(3) To what extent do customers perceive there to be viable alternative modes for a given shopping occasion? Previous studies have largely neglected this question, implicitly or explicitly assuming that each shopping activity involves a true choice among competing alternatives. It is obviously important to test that assumption, and identify variables associated with perceived mode captivity versus choice.

(4) Are the various shopping modes substitutes, or complements? For example, do people who do a lot of e-shopping tend to do less store shopping, or do high amounts of one tend to be associated with high amounts of the other? Do former catalog shoppers replace the catalog with the internet, or supplement it; conversely, do internet shoppers become new catalog shoppers as retailers engage in cross-channel marketing? Are there different relationships for different segments of the population? For example, shopping modes may be complementary for innate “shopaholics”, but substitutionary for the time-pressured. Although to keep the survey at a reasonable length it is not possible to obtain the data for a rigorous analysis of transportation impacts, the answers to these questions will have direct implications for the likely transportation impacts of e-shopping.

The remainder of this report focuses on the administration of a survey instrument that will facilitate the research directions discussed above. The organization of the document is as follows. Section 2 discusses the design of the sampling plan for the data collection. The following section then describes the actual sampling process. The fourth section briefly presents the survey design. Section 5 discusses the development of the survey instruments and Section 6 outlines the deployment of the survey. Next, a section is devoted to the steps taken to build the final sample through data cleaning. A concluding section ends the report.

2. SAMPLING PLAN

Obtaining a strictly representative sample is not essential to the success of this study. This is because our purpose is not primarily to report descriptive statistics for the sample distributions of various measures of interest and expect them to faithfully represent their population counterparts, but rather to identify relationships among the variables we measure. Those relationships can be reliably captured, even when raw univariate distributions are not representative. For example, the sample may over-represent some income groups and under-represent others, but (assuming all income groups have a numerically substantial representation in the sample) the distribution of other responses given a certain income level can still be properly represented (see Brownstone, 1998 for a discussion of this issue in the discrete choice modeling context, and Babbie, 1998 for a general comment).

What is critical, then, is to have sufficient diversity with respect to the variables of interest. In particular, it is important to have a substantial number of e-shopping occasions in the sample. At the same time, there is currently not a convenient way to systematically sample internet users, let alone e-shoppers. Thus, we conducted a regular mail recruitment of the sample, but for a setting in which a random sample of residents is expected to net a high level of internet literacy and shopping activity. Specifically, the sampling plan segments our population by city, neighborhood type, and census tract. With respect to city, we split the sample between our university town of Davis, California (pop. 60,000), which is on the fringe of the medium-size Sacramento Metropolitan Statistical Area (MSA; pop. about 2 million) but separated from the rest of the area by agricultural land and flood control plains, and Santa Clara (pop. 100,000), an affluent, computer-literate community in Silicon Valley (embedded within the large San Francisco MSA, pop. about 7 million) that also contains a private university (enrollment 8,000). Figure 1 portrays the two study areas from about 45,000 feet up. Although the college-town sample will contain a higher-than-average proportion of younger adults, that actually has some advantages, since those respondents will tend to be harbingers of future adoption patterns. That is, the ICT behavior of today's over-50 adults is apt to tell us less about the future than that of today's 20-somethings. Together, the two subsamples will allow for some limited testing of the role of urban context in shopping mode choice, a factor that has been found important in the research of Farag, et al. (2006) and Krizek et al. (2005).

Within those two cities, we targeted "traditional" neighborhoods, meaning those with closely-spaced perpendicular streets and nearby retail stores, and "suburban" neighborhoods, which are typical of more recent developments and contain curved streets, cul-de-sacs, and major-intersection-located retail strip malls. Though both Davis and Santa Clara are largely suburban communities and such segmentation may, in the end, be somewhat arbitrary, we wanted to have at least the possibility of measuring the impact of different neighborhood types within the two cities.

Due to budget constraints, a recruitment sample size of 8,000 was established with the hope (based on the prior experience of the second author with several paper surveys of similar complexity, albeit shorter than this one) of receiving on the order of one to two thousand valid responses. The sample was first distributed evenly among city and neighborhood type (see Table 1), targeting 2,000 residents in each neighborhood type within each city. To operationalize the administration of the survey, census tracts within each desired city and neighborhood were selected through map inspection and site visits. Within each neighborhood type, the number of households in each targeted census tract, per the 2000 census, was used to allocate the sample by tract. As shown in Table 1, the sampling rate for each tract ranged from 0.382 to 0.582.

As discussed previously, the goal of the sampling plan was not to represent the population of the United States, or any other general population for that matter. Rather, given limited resources, the goal was to sample a few typical areas offering a diversity of residential neighborhood type and urban context, and use the data to investigate a variety of variables' s on shopping and travel behavior. Table 2 presents a brief summary of characteristics for the targeted census tracts. We can see that the tracts have a rather wide range of median incomes, structure age, and owner/renter mix. It should be mentioned that while the median incomes in the "traditional" Davis neighborhoods appear to be rather low, many residents in these areas are university students who may have spending powers beyond what their nominal income would suggest.

Figure 1: Aerial Photos of Davis (top) and Santa Clara (bottom), California

Table 1: Target Sampling Plan

City	Neighborhood Type	Census Tract	Households	N			Sampling Rate
Santa Clara	Suburb	5050.01	2,465	4,000	2,000	1,432	0.581
		5050.07	976			568	0.582
	Traditional	5056	1,246		2,000	708	0.568
		5057	2,271			1,292	0.569
Davis	Suburb	105.07	3,220	4,000	2,000	1,231	0.382
		105.08	1,055			403	0.382
		106.05	956			366	0.383
	Traditional	106.02	2,393		2,000	1,058	0.442
		107.01	2,129			942	0.442

Table 2: Media Income, Structure Age, and Tenure by Census Tract

City	Neighborhood Type	Census Tract	Median Income (1999)	Structure Age		Tenure	
				Before 1970	After 1970	Owner	Renter
Santa Clara	Suburb	5050.01	$92,613	19.5%	80.5%	50.2%	49.8%
		5050.07	$74,911	76.8%	23.2%	66.2%	33.8%
	Traditional	5056	$42,625	56.2%	43.8%	22.4%	77.6%
		5057	$53,657	80.8%	19.2%	31.1%	68.9%
Davis	Suburb	105.07	$42,813	10.1%	89.9%	42.4%	57.6%
		105.08	$62,537	8.3%	91.7%	61.6%	38.4%
		106.05	$80,238	0.0%	100.0%	78.1%	21.9%
	Traditional	106.02	$25,803	47.0%	53.0%	18.9%	81.1%
		107.01	$24,204	74.8%	25.2%	24.8%	75.2%

3. SAMPLING

To implement the sampling plan, 15,000 names and addresses were purchased from the firm Martin Worldwide, Incorporated (http://www.martinworldwide.net/). We requested that the addresses be located as summarized in Table 1. Martin Worldwide indicated the request could be satisfied, with a total of 7,500 (rather than 4,000) addresses to be delivered in each city (each tract's sample size is then increased proportionally to the number of households in the tract – see Table 3).

After receiving the list of 15,000 names and addresses from the vendor, the list was pruned as follows. First, because we wanted to locate each address in one of the target census tracts, we removed post office box addresses from the list (1,484 of the 15,000). Next, duplicate names, with differing addresses, were removed (1,023). The remaining addresses were then geo-coded using the website http://www.batchgeocode.com/. This website allows for a batch of addresses to be entered and plotted on an interactive map. Twenty-two of the remaining addresses could not be geo-coded. The resulting number of "good" addresses was 12,471.

Using geographic information system (GIS) software, the remaining 12,471 were mapped against the target census tracts. Figure 2 presents the shaded census tracts along with dots representing each of the addresses for Santa Clara (on the left) and Davis (on the right). As the figure demonstrates, the addresses are in the same general area as the census tracts, but they are not entirely contained within them.

The graphical data of Figure 2 are summarized in Table 3. For Santa Clara, Martin Worldwide promised to deliver 7,500 addresses in the four target census tracts and delivered 1,398 on-target addresses (well short of our requested 4,000). In Davis, 2,057 on-target addresses were delivered.

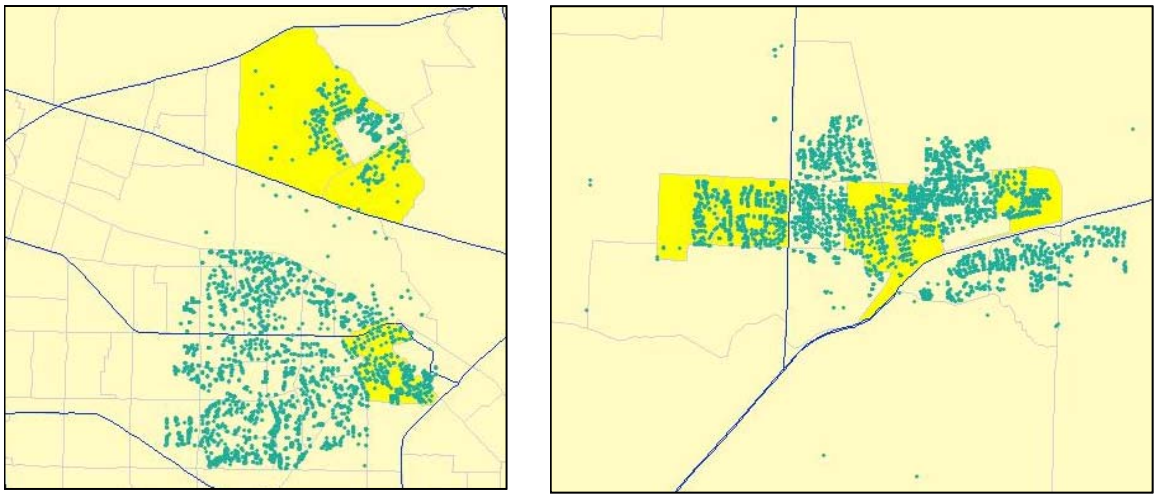


Figure 2: Address Locations and Desired Tracts

Combining the "on-target" addresses from Santa Clara and Davis, a total of 3,455 addresses in the desired census tracts were delivered by Martin Worldwide. Not wanting to purchase more addresses (which could also be off-target) from another vendor and recognizing that the differences between adjacent neighborhoods within Santa Clara and Davis may not generally be dramatic, it was decided to build a sample of 8,000 addresses from those provided by Martin Worldwide.

Table 3: Delivered Addresses by Census Tract

Location			Promised N	Requested N	Delivered N	
City	Neighborhood	Tract			On-Target	Off-Target
Santa Clara	Suburb	5050.01	2,686	1,432	542	---
		5050.07	1,064	568	232	---
	Traditional	5056	1,328	708	165	---
		5057	2,422	1,292	459	---
	Total		7,500	4,000	1,398	5,267
Davis	Suburb	105.07	2,308	1,231	663	---
		105.08	756	403	253	---
		106.05	686	366	370	---
	Traditional	106.02	1,984	1,058	409	---
		107.01	1,766	942	362	---
	Total		7,500	4,000	2,057	3,749

The sample was built by first measuring the distance from each address to the boundary of each census tract (using GIS software). Next, the points closest to the census tracts were included in the sample. This was done sequentially, starting with the traditional tracts, taken as a group, and then the suburban tracts. For example, in Santa Clara we started with the 624 (165 + 459) on-target addresses in the traditional tracts. We then took the next 1,376 addresses closest to either of the traditional tracts until we reached our "quota" of 2,000 addresses. We then took the 774 (232 + 542) on-target addresses for the suburban tracts, and added the 1,226 addresses closest to them. This process was repeated for the Davis addresses. With the realization that some of the more distant addresses could potentially be assigned to either neighborhood type, assignment to the traditional tracts was given priority under the assumption that in both Davis and Santa Clara, traditional neighborhoods are scarcer than suburban ones.

Table 4 summarizes a variety of distance measures for the off-target addresses (the "on-target" addresses are not included in the average distance calculations). The table shows that the median distance from the tracts to the addresses is, in three of four cases, less than half a kilometer. The Davis addresses are, in general, closer to their targets than the Santa Clara addresses.

4. SURVEY DESIGN

As discussed in the next section, both web- and paper-based survey instruments were developed as part of the study. A brief description of the survey contents is presented here.

The survey started with a simple *Welcome* question, intended to be easy and fun to answer: "If you HAD to spend an hour or two shopping, where would you prefer to be?" Response options were: downtown shopping district, shopping mall, bookstore, grocery store, electronics store, hardware/home improvement store, and "other (please specify)".

Part A: Shopping and General Shopping-related Attitudes followed the *Welcome* section and contained 42 statements respondents were able to agree or disagree with on a five-point Likert-type scale ("strongly disagree", "disagree", "neutral", "agree", "strongly agree"). In the design stage, we identified 21 attitudinal dimensions of interest, based on a thorough review of the literature (Cao and Mokhtarian, 2005) and our own judgment. Then, drawing from previously published surveys and our own judgment, we prepared the following 79 statements in the 21 categories:

Table 4: Distance Measures for Off-Target Addresses

Measure		Santa Clara		Davis	
		Traditional	Suburban	Traditional	Suburban
On-target addresses		624	775	771	1,286
Off-target addresses		1,376	1,225	1,229	714
Distance from selected tracts for off-target addresses (meters)	Minimum	0.2	6.0	0.1	0.9
	Maximum	1,435.7	3,344.7	679.6	431.3
	Mean	715.8	1,132.0	335.5	255.7
	Standard deviation	446.1	1,148.3	208.8	139.1
	Median	665.0	277.1	354.2	304.6

- *Technology – General*
  - I am often among the first to buy new technological products.
  - I often find high-tech products difficult to operate.
  - I like to track the development of new technology.
  - Technology brings at least as many problems as it does solutions.
  - On the whole, technology makes our lives better.
- *Technology – Computers*
  - I enjoy using computers.
  - The internet makes life more interesting.
  - Computers are more frustrating than they are fun.
  - The internet plays only a minor role in my life.
- *Variety seeking – General*
  - "Variety is the spice of life."
  - I like a routine.
  - Change is unsettling.
  - Change is refreshing.
- *General innovativeness – General*
  - I am generally cautious about accepting new ideas.
  - I prefer to see other people using new products before I consider getting them myself.
  - I like to experiment with new ways of doing things.

- *eShopping innovativeness – Shopping specific*
  - I know more about shopping over the internet than most of my friends.
  - I am among the last in my circle of friends to purchase something over the internet.
  - I enjoy taking chances in buying over the internet just to enrich my shopping experience.
- *Price consciousness – Shopping specific*
  - If I really want something, I'll often buy it even if it costs more than it should.
  - I generally compare prices before buying.
  - I look for sales and special offers when I'm shopping.
  - Sales and special offers encourage unnecessary spending.
  - Generic brands usually offer just as good a quality as more expensive ones.
  - It's important to me to get the lowest prices when I buy things.
- *Time consciousness – Shopping specific*
  - I'm too busy to shop as often or as long as I'd like.
  - Being a smart shopper is worth the extra time it takes.
  - I'm often in a hurry to be somewhere else when I'm shopping.
  - Saving time when I shop is important to me.
- *Impulsive buying – Shopping specific*
  - I generally stick to my shopping lists.
  - Before buying something, I generally take some time to think it over.
  - I often make unplanned purchases.
- *Shopping enjoyment – Shopping specific*
  - For me, shopping can be an important leisure activity.
  - I would often prefer someone else to do my shopping.
  - Under the right circumstances, shopping is fun.
  - I enjoy the social aspects of shopping.
  - Shopping is usually a chore for me.
  - Shopping helps me relax.
- *Risk aversion – General*
  - Taking risks fits my personality.
  - Before I make a decision, I need to be sure it's the right one.
  - I'm willing to take a big chance for the possibility of a big payoff.
- *Trustingness – General*
  - Most people can be counted on to do what they say they will do.
  - One should be very cautious with strangers.
  - People are generally trustworthy.

- *Attitude towards credit cards – General*
  - I like to pay for everything with cash.
  - Credit cards encourage you to spend unnecessarily.
  - I need to have a credit card.
- *Travel – Shopping specific*
  - Shopping is too physically tiring to be enjoyable.
  - Going shopping is mentally stressful.
  - I like to stroll through shopping areas.
  - Getting to where I usually shop is a hassle.
  - Sometimes for me, shopping is mostly an excuse to get out of the house or workplace.
  - Even if I don't end up buying anything, I still enjoy going to stores and browsing.
- *Travel – General*
  - "Getting there is half the fun".
  - The only good thing about traveling is getting to the destination.
  - Travel time is generally wasted time.
  - In my daily travel, I try to make good use of my time.
  - The routine traveling that I need to do interferes with doing other things I like.
- *Physical exercise – General*
  - Whenever possible, I prefer to walk or bike rather than drive.
  - I follow a regular physical exercise routine.
  - I should be more physically active than I am.
- *Environmental concerns – General*
  - To improve air quality, I am willing to pay a little more to use a hybrid or other clean-fuel vehicle.
  - We should raise the price of gasoline to reduce congestion and air pollution.
  - We need more public transportation, even if taxes have to pay a lot of the costs.
- *Environmental concerns – Shopping specific*
  - Shopping travel does not cause very much air pollution.
  - We should try to reduce shopping travel to benefit the environment.
  - A lot of the packaging used for products is wasteful.
- *Status – General*
  - For me, a lot of the fun of having something nice is showing it off.
  - I am often the one introducing a new trend to my friends.
  - Impressing other people is of little interest to me.

- *Materialism – General*
  - My lifestyle is relatively simple, in terms of material goods.
  - I would/do enjoy having a lot of expensive things.
  - As long as I have the basics, material things don't matter to me that much.
- *Spending money enjoyment – General*
  - I enjoy spending money.
  - Buying things cheers me up.
  - If I got a lot of money unexpectedly, I would probably save more of it than I spent.
- *Social influence*
  - My friends and I share many common interests.
  - When my friends try something new, I like to try it as well.
  - A lot of the fun of trying new things is sharing the experience with friends.

For the survey, the 21 categories were combined and eliminated to form 19 categories, and 42 questions were selected and modified. In keeping with guidance from the survey design literature (e.g. Baumgartner and Steenkamp, 2001; Ellard and Rogers, 1993), we diversified the directionality of the final list of statements, to reduce the tendency to fall into an automatic response mode. We made an effort to include at least one positively-oriented and one negatively-oriented statement for each construct, but where we could not produce satisfactory statements by that guideline, we did not force it. The factor analysis literature (e.g. Fabrigar *et al*., 1999) further advises including 3-5 items (statements) for each hypothesized construct, but in view of the large number of constructs we considered important to our context, and the interconnectedness of many of them (thus leading to their merging or overlapping in an exploratory factor analysis), we limited the number of statements per construct to two in most cases – again as a design compromise to reduce respondent fatigue. The final list, sorted alphabetically by category label, is included in Table 5 and Table 6.

Part B: Your Purchasing Experiences asked if the respondent had purchased items in 15 product classes over the past year via the internet, in a store, through a catalog, or by other means. This section was originally more elaborate, asking for the frequency (category) of purchasing each product class by each mode, but was simplified to reduce the burden on the respondent without entirely sacrificing information on shopping patterns for items other than the two main product classes of interest.

Part C: A Recent Purchase asked a series of detailed questions about a specific recent purchase in one of two product classes (the "search" goods of books/CDs/DVDs/videotapes, or the "experience" goods of clothing/ shoes), and were purchased via one of three modes: over the internet, in a store, or through a catalog. The questions obtained situational information about the specific purchase occasion, and also established the means by which the respondent initially became aware of the item, obtained information about it, and experienced the item before purchasing it.

In *Part D: Advantages and Disadvantages of Different Ways of Shopping*, respondents were asked to imagine that they will soon be making a purchase similar to the one discussed in Part C. They were then invited to evaluate two of the three shopping modes – store, internet, and catalog – with respect to such a purchase. The decision to present only two modes was again a design concession to reduce respondent fatigue. The evaluation consisted of 28 statements (for each mode) that the respondents were again asked to agree or disagree with on a five-point, Likert-type scale. The first set of statements related to the store mode for all respondents, as the "anchor" with which it was presumed they would all be familiar. Since

the catalog mode was of secondary interest to the study (in view of the necessity of reducing the respondent burden), it was presented in the second set of parallel statements only if it were the chosen mode for the key purchase; otherwise the second mode was the internet.

Similar to the process for the Part A attitudinal questions, in the design stage, we started with 50 statements in 14 categories. In Table 7 and Table 8, these statements are shown as specific to the internet; in the survey administration, minor changes were made to the statements to make them specific to store and catalog shopping in turn. In the final survey, these statements were reduced to 28 in 13 categories, as shown in Table 9 for the store mode.

Table 5: General Shopping-Related Attitudinal Statements, First Nine Categories

Hypothesized Construct	*	Survey Statement
Credit cards	–	Credit cards encourage unnecessary spending.
	–	I prefer to pay for things by cash rather than credit card.
Environmental–general	+	We should raise the price of gasoline to reduce congestion and air pollution.
	+	To improve air quality, I am willing to pay a little more to use a hybrid or other clean-fuel vehicle.
Environmental–shopping-related	–	Shopping travel creates only a negligible amount of pollution.
	+	A lot of product packaging is wasteful.
Exercise	+	Whenever possible, I prefer to walk or bike rather than drive.
	+	I follow a regular physical exercise routine.
Impulse buying	+	When it comes to buying things, I'm pretty spontaneous.
	–	I generally stick to my shopping lists.
Innovation	–	I am generally cautious about accepting new ideas.
	–	I prefer to see other people using new products before I consider getting them myself.
Materialism	+	I would/do enjoy having a lot of expensive things.
	–	My lifestyle is relatively simple, in terms of material goods.
Price conscious	–	It’s too much trouble to find or take advantage of sales and special offers.
	+	It’s important to me to get the lowest prices when I buy things.
Risk-taking	+	Taking risks fits my personality.
	–	“Better safe than sorry” describes my decision-making style.

* Directionality with respect to construct label.

Table 6: General Shopping-Related Attitudinal Statements, Final Ten Categories

Hypothesized Construct	*	Survey Statement
Shopping enjoyment	–	Shopping is usually a chore for me.
	+	I enjoy the social interactions shopping provides.
	+	Shopping helps me relax.
	+	Shopping is fun.
Spending money	+	If I got a lot of money unexpectedly, I would probably spend more of it than I saved.
	+	Buying things cheers me up.
Status	+	I often introduce new trends to my friends.
	+	For me, a lot of the fun of having something nice is showing it off.
Technology–computer-related	+	The internet makes my life more interesting.
	–	Computers are more frustrating than they are fun.
Technology–general	+	I like to track the development of new technology.
	–	Technology brings at least as many problems as it does solutions.
Time consciousness	+	I'm often in a hurry to be somewhere else when I'm shopping.
	+	I'm too busy to shop as often or as long as I'd like.
Travel–general	+	I am generally doing productive or enjoyable things, such as making phone calls or listening to the radio, while traveling.
	–	The only good thing about traveling is getting to the destination.
Travel–shopping-related	+	Even if I don't end up buying anything, I still enjoy going to stores and browsing.
	–	Shopping is too physically tiring to be enjoyable.
	+	I like to stroll through shopping areas.
	+	For me, shopping is sometimes an excuse to get out of the house or workplace.
Trust	+	People are generally trustworthy.
	–	I tend to be cautious with strangers.
Variety-seeking	–	I like a routine.
	+	"Variety is the spice of life."

* Directionality with respect to construct label.

Table 7: Design-stage Mode-Specific Attitudinal Statements, First Five Categories (Internet Mode)

Category	Statement
Customer service	When shopping over the internet, I am confident of getting a desired item within a reasonable amount of time.
	If necessary, it is easy to return a product purchased over the internet.
	Internet shopping provides poor after-purchase customer service.
	Internet retailers are generally very receptive to customer feedback.
Trust	In my experience, most internet stores keep their commitments.
	I am concerned that internet stores will fail to meet my expectations.
	I am confident in my ability to determine whether a retailer is trustworthy.
	I prefer to shop the internet sites of national chain stores.
	I value the anonymity that shopping on the internet provides.
Ease of use	Internet shopping is easy.
	The product information I need is easy to find over the internet.
	I often find shopping over the internet to be frustrating.
Availability/Selection	Shopping over the internet makes it easier to obtain certain products that are hard to find elsewhere.
	A lot of times, products I want are unavailable over the internet.
	When it comes to [clothing, books], I can find anything I want for sale over the internet.
	Certain products I purchase are only available over the internet.
Search costs (effort savings)	The internet makes it easy to check the availability of products.
	The internet makes it simple to compare products.
	It is difficult to compare products over the internet.
	It is easy to get information from a live person when shopping over the internet.
	It takes too long to obtain product information over the internet.

Table 8: Design-stage Mode-Specific Attitudinal Statements, Last Nine Categories (Internet Mode)

Category	Statement
Time savings	All things considered, buying over the internet saves me time.
	The internet sites I use allow me to fulfill many of my shopping needs in just one location.
Gratification delay	When shopping over the internet, I am confident of getting a desired item within a reasonable amount of time.
	I often have to wait too long to receive a product purchased over the internet.
Money savings	All things considered, buying over the internet saves me money.
	Considering shipping costs, [clothing,books] are usually more expensive when purchased over the internet.
Information (broad, fast, comparison)	Internet stores often fail to offer enough product information.
	It takes too long to find a desired product on the internet.
	The product information provided by internet stores is generally up to date.
	Product information on the internet is clear and understandable.
Convenience	The internet allows me to shop at any time I wish.
	Internet shopping is available any time I want it.
	Internet shopping is available to me anywhere I would like it to be.
	I enjoy being able to shop from home without having to get dressed and go out.
	Having to get dressed and go to the store is a hassle.
	The stores I want are conveniently located.
Product risk	Internet stores often provide misleading product information.
	Internet shopping generally enables me to experience products before buying to the extent that I want to.
	Products purchased over the internet often fail to meet my expectations.
	I'm concerned that a product I purchase over the internet will not perform as expected.
Financial risk	It is risky to release credit card information over the internet.
	I am generally nervous about providing personal information over the internet.
	Potentially having to pay a fee to return an unsatisfactory product is a reasonable risk to take.
	The prospect of having to return a product that I've purchased over the internet doesn't really bother me.
	I'm concerned that an internet store will fail to deliver a product I've purchased.
General enjoyment	I enjoy shopping over the internet.
	Shopping over the internet is boring.
Store-brand attachment	When it comes to [clothing, books], I have a strong preference for shopping at one or a few particular internet sites.
	With respect to buying [clothes, books], I am always on the lookout for a new internet site to check out.

Table 9: Final Mode-Specific Attitudinal Statements (Store Mode)

Category	*	Statement
Availability/ selection	+	When it comes to buying books/CDs/DVDs/ videotapes, I can find anything I want in stores.
	−	A lot of times, products I want are unavailable in stores.
Convenience	+	The stores I want/need to shop at are conveniently located.
	+	Getting dressed and going out is an enjoyable aspect of store shopping for me.
	+	Stores are open whenever I want to shop.
Customer service	−	Stores typically provide poor after-purchase customer service.
	+	If necessary, it is easy to return a product purchased at a store.
Ease of use	+	The product information I need is easy to find in stores.
	−	I often find shopping in stores to be frustrating.
Financial risk	+	It is risky to release credit card information to stores.
	+	I am uncomfortable about providing personal information to stores.
General enjoyment	−	Shopping in stores is boring.
	+	I enjoy shopping in stores.
Gratification delay	+	I often have to wait too long for a store to obtain the product I want to purchase.
	−	When shopping in stores, I am able to immediately obtain the products I purchase.
Money savings	−	Considering taxes and other costs, books/CDs/ DVDs/videotapes are usually more expensive when purchased in stores.
	+	All things considered, buying in stores saves me money.
Product risk	+	I'm concerned that a product I purchase in a store will not perform as expected (e.g. quality, etc.).
	−	When shopping in stores, I am able to experience products before buying, to the extent that I want to.
Search costs (effort savings)	−	It is difficult to compare products at stores.
	+	When shopping in stores, it is easy to check the availability of products.
Store-brand attachment	−	With respect to buying books/CDs/DVDs/ videotapes, I am always on the lookout for a new store to check out.
	+	When it comes to books/CDs/DVDs/ videotapes, I have a strong preference for shopping at one or a few particular stores.
Time savings	+	I value stores that allow me to fulfill many of my shopping needs in just one location.
	+	All things considered, buying in stores saves me time.
Trust	+	I prefer to shop at independent stores rather than national chains.
	−	I value the anonymity (e.g. paying with cash) that shopping in stores provides.
	−	I am concerned that unfamiliar stores will fail to meet my expectations.

* Directionality with respect to construct label.

In *Part E: Frequency of Shopping*, more general questions are asked about the frequency of shopping by mode for the key item discussed in Parts C and D. *Part F: Your Use of Internet and Communication Technology* continues the move back to the general by asking questions about internet use, as well as the use of other technologies. Finally, *Part G: Some Information About Yourself* asks general sociodemographic questions that will allow our sample to be compared to more general populations.

5. SURVEY INSTRUMENT DEVELOPMENT

The survey was administered primarily over the internet, though three separate paper survey versions were also developed. This section discusses the development of both instrument types.

The internet version of the survey was developed using the commercial software vendor Zoomerang (http://www.zoomerang.com). Zoomerang allows users to develop surveys and then handles the administration and data collection of the surveys. Though the software does limit the question types and format of the instrument, the advantages in terms of security and data collection made the choice to use a commercial vendor superior to collecting the data on our own. The advantages and disadvantages, in the context of this survey, of using a web-based commercial vendor in general, and Zoomerang in particular, are as follows:

- *Advantages*
  - Data-entry is done automatically and accurately;
  - Issues related to web-security and server/database management are taken care of;
  - Development time is much faster;
  - Data can easily be downloaded;
  - Relatively low cost;
  - The software is constantly improving.
- *Disadvantages*
  - The types of questions that can be asked are limited;
  - Branching logic can only be related to a single response at a time;
  - The software did not permit the respondent to save a partially-completed survey and automatically return to the same point at the next log-in;
  - Hitting the “back” button on one’s web browser and changing one’s response on a previous screen did not overwrite the previous response, so the only way respondents could change their answers after moving past a page was to start over entirely;
  - Only one type of “button” is available for each question type and in the case of multiple choice questions, the difference between the “clicked” and “not clicked” images is not great;
  - The font size and style can be changed using html tags, but not on a global level (i.e. each statement/question/instruction required its own html tag);

- Entered data, such as an identification number, cannot be checked against an existing database;

- Although individual responses could be limited to the provided options, multiple responses could not be checked for internal consistency (e.g. that number of workers did not exceed number of household members) in real time;

- No "thermometer" showing percent of survey completed was automatically available; we manually created one at several points in the survey;

- The order of questions could not be made random;

- Respondents could not be automatically prompted to complete missing responses without forcing them to do so – i.e. there was no ability just to remind respondents that a field was left blank in case it was inadvertent, yet to reduce respondent irritation we wanted to limit the number of questions for which we required a response;

- Neither survey pages nor questions could be copied and pasted, which made creating this survey incredibly tedious and subject to typographical errors;

- Zoomerang customer service was largely unresponsive to problems with their software, which, as with all software, did have bugs;

- The survey is administered on a Zoomerang-hosted website, which made a few respondents concerned about the legitimacy of our claimed relationship with UC Davis.

The Zoomerang service works by allowing the user to design any number of survey pages. Within each survey page the user can include any number of survey questions. There are a variety of question types, including multiple choice, ranking, write-in, etc. Answers to any particular question can then be used to facilitate skipping logic between pages (e.g. if the answer to Question 4 is "yes", go to page 4). After developing the survey, a website is then built for survey administration. The Zoomerang software then takes care of the administration and data storage.

Figure 3 presents a flow chart of the web pages used in the survey administration and this figure is discussed in detail here. Each box in the figure represents a group of web pages. The top number in the box refers to the web page numbers included in that section; the next line gives the "Part" of the survey of which the pages belong to (see Section 4); the last line gives a description of the pages.

The first page in the survey and the top box in Figure 3 requires the user to enter a 10-digit identification number (see Section 6 of this report for more information on the ID number). The next three boxes in the figure reference the Welcome, Part A, and Part B portions of the survey, which include web pages 2, 3 to 5, and 6, respectively. All users complete these portions of the survey.

The branching starts in Part C, which focuses on the purchase of a particular item in one of two product classes – books/ CDs/DVDs/videotapes ("search" goods) or clothing/shoes ("experience" goods) – purchased via one of three modes: over the internet, in a store, or through a catalog. We narrowed the questioning to specific product classes on the assumption (supported by other research) that relevant variables could be weighted differently depending on the nature of the product. We chose relatively low-cost, high-frequency-of-purchase product types to ensure the presence of a sufficient proportion of relatively recent purchase occasions in the sample, and we chose these two to represent the difference between experience goods (those often needing to be tried in some way before being purchased) and

search goods (those that can often be satisfactorily evaluated on the basis of externally-provided information alone; Peterson et al., 1997). The product class pertaining to a given respondent will henceforth be referred to as the "key purchase" or "key item".

The first branching is based on the purchase mode. We first inquire about a recent (within six months or so) internet purchase of one of the key items (on page 7). If such a purchase was made, the respondent is directed to page 8 and the survey continues with more detailed questions about the internet purchase. If such a purchase was not made, we inquire about a recent purchase made in a store (page 9). Again, if such a purchase has been made, the survey continues with more detailed questions about the store purchase. If no recent store purchase has been made, we ask about a catalog purchase. If no *recent* purchase of a key item has been made via any of these three modes, the survey inquires about *any* purchase of a key item the respondent can recall. If the respondent cannot recall the purchase of any of the key items (book, CD, DVD, videotape, clothing, or shoes), they are directed to Part F of the survey.

The order of the mode questions heavily influences the "mode shares" of the key purchases. As such, we made two versions of the internet survey. The first inquired about internet purchases first (and store purchases second), as presented above and as shown in the figure. The second asked about store purchases first (and internet purchases second). The two surveys were deployed at different times (out of view of the respondents) to try and gather a balanced sample in terms of store-versus-internet mode shares. The internet-first version was active from June 1 - 13, 2006 and collected 439 (not all unique or complete) responses; the store-first version was active from June 14 - September 14, 2006 and collected 649 responses (again, not all unique or complete). As the catalog aspect of the survey was of secondary interest, the catalog option was always asked last.

After branching based on shopping mode, Part C then branched on item type. Pages 8, 10, 12, and 14 to 16 ask which key item, among books, CDs, DVDs, videotapes (the search goods) and clothing and shoes (experience goods) was purchased most recently. This branching resulted in respondents being directed down one of six Part C tracks representing the item-mode combination of their key purchase, namely: book-internet (pages 17 to 27; "book" is used to represent all the search goods), clothing-internet (pages 28 to 38; "clothing" is used to represent all the experience goods), book-store (39 to 49), clothing-store (50 to 60), book-catalog (61 to 71), and clothing-catalog (72 to 82). In this portion of the survey, the respondent answers a series of detailed questions regarding the key purchase.

Part D of the survey asks respondents to compare shopping modes in the context of purchasing the key item they described in Part C. First, each respondent is asked to respond to 28 statements about purchasing the key item, either for books/CDs/DVDs/videotapes or clothing/shoes, in a store. Next, the respondent does the same for 28 companion statements for either the internet (if the item was purchased in the internet or store) or a catalog (if the key item was purchased in a catalog). Thus, Part D collapses into the four tracks (book-store + book-internet; clothing-store + clothing-internet; book-store + book-catalog; clothing-store + clothing-catalog) shown in Figure 3.

Part E asks only item-specific questions, thus collapsing the tracks into two, book and clothing. Finally, Parts F and G are asked of each respondent.

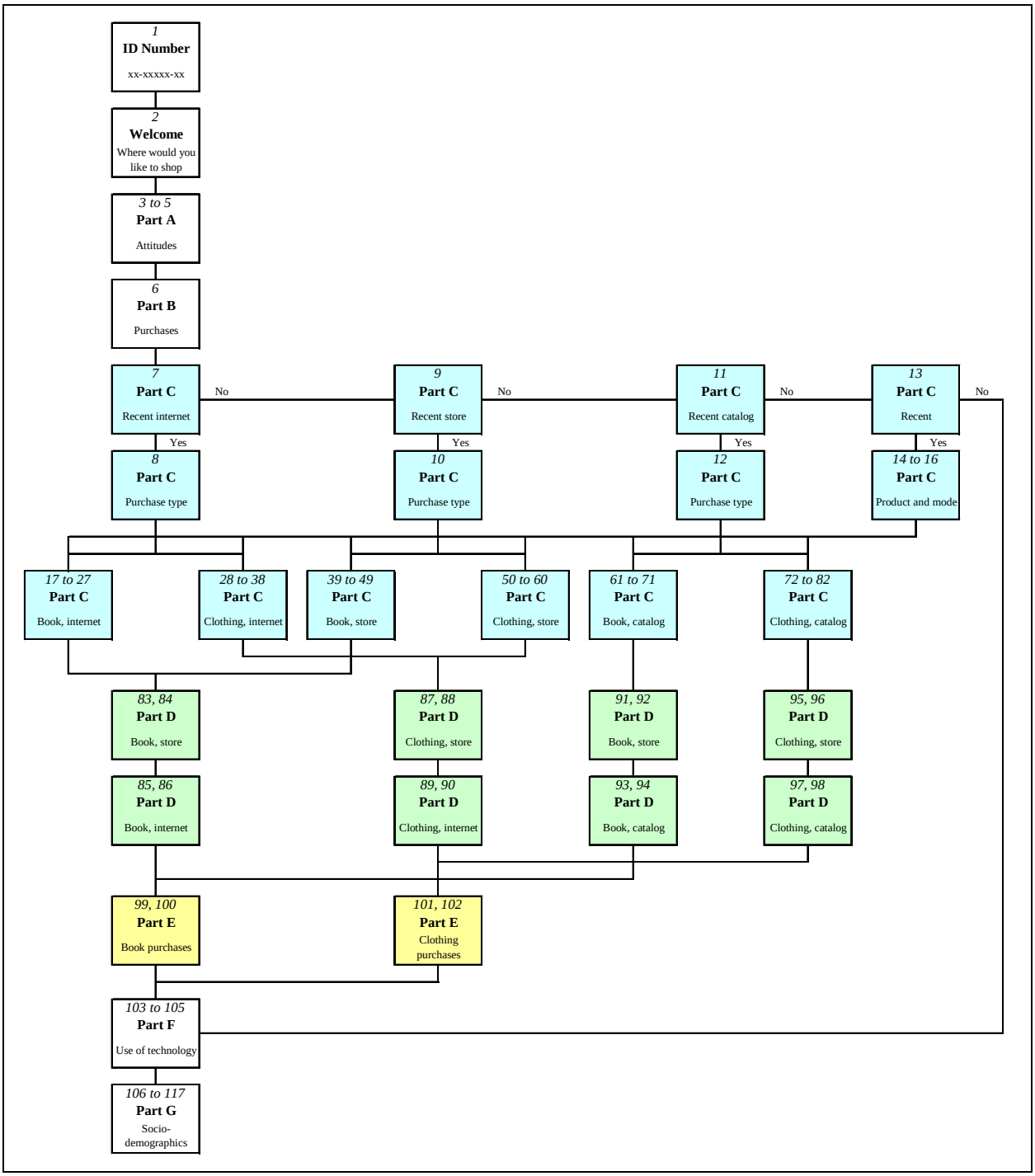


Figure 3: Internet Instrument Flow Chart

As discussed in the next section, paper versions of the survey were made available to those with either a preference for a paper survey or an inability to complete the internet version. Because of the limiting nature of a paper survey, the full branching options included in the web survey were not available to those

completing the paper surveys. To partially reduce this constraint, the following three paper surveys were administered:

- Version 1: Book
  - Captures book/CD/DVD/videotape purchases;
  - Part C: asks first about a recent internet purchase; if none is recalled, asks about a store purchase;
  - Part D: first set of statements relates to store purchases, second set to internet purchases.
- Version 2: Clothing-internet
  - Captures clothing/shoe purchases;
  - Part C: asks first about a recent internet purchase; if none is recalled, asks about a store purchase;
  - Part D: first set of statements relates to store purchases, second set to internet purchases.
- Version 3: Clothing-catalog
  - Captures clothing/shoe purchases;
  - Part C: asks first about a recent catalog purchase; if none is recalled, asks about a store purchase;
  - Part D: first set of statements relates to store purchases, second set to catalog purchases.

As those requesting a paper survey had to do so by e-mail or phone, screening questions were used to determine the most appropriate paper survey for each user. The screening questions are as follows:

1. Do you have access to the internet? Yes → Question 2; No → Version 3;
2. (If "yes" to question 1:) Do you shop more often for books or clothing? Books → Version 1; Clothing → Version 2.

It should be noted that Versions 1 and 2 (and the catalog track of Version 3) of the paper surveys are subsets of the web-based survey and, as such, the results of the paper survey were entered directly into the web-based survey. A modified web survey was developed for the store track of Version 3 in order to enter those surveys into the database. A different survey was needed because the Version 3 paper surveys presented catalog- and store-specific Part D statements, whereas in the web survey, those purchasing items in a store are presented store- and internet-specific Part D statements. The modified web survey included the catalog- and store-specific Part D statements for those making a purchase in a store.

6. DATA COLLECTION

The data were collected over a three-month period from June to August, 2006. The first step in the process was sending out a recruitment letter to each of the 8,000 selected addresses. We debated several different ways to address the envelopes and letters, either to the householder by name (as provided by the vendor), to "Current Resident", or to "[name] or Current Resident", and whether to treat the envelopes the

same way as the letters. The advantage of the latter two approaches for the envelope is that no mail should be returned as undeliverable due to the occupant in the vendor's database having moved; the obvious disadvantage is that the letter is immediately marked as a mass mailing, possibly "junk mail", and therefore more likely to be discarded. On the other hand, using the resident's name only was potentially "friendlier", but also potentially more threatening or annoying to some ("how did you get my name and know where I live?"), and had the additional disadvantage that if the resident had moved, the letter would be returned as undeliverable (a non-trivial factor, as shown below). Ultimately, we chose to use the resident's name on the envelope, and "Dear [city] resident" as the salutation on the letter – hoping that the name on the envelope would get it opened, while the anonymity of the salutation would reduce the threat level. The envelopes were especially printed with the return address of "Prof. Patricia L. Mokhtarian", to pique curiosity and to further distinguish the recruitment from commercial solicitations.

The recruitment letters were identical for Davis and Santa Clara residents, except for (1) the two appearances of the city name, (2) the use of two different telephone numbers for David Ory so that calls to him would be essentially local for residents of either city (a 530 area code number for the Davis letters, and a 415 area code number for the Santa Clara letters), and (3) the word "(collect)" after the 530 area code telephone number for Patricia Mokhtarian on the Santa Clara letter. A copy of the letter sent to the Davis residents is presented in Figure 4.

The letters were printed by the Reprographics Division of the University of California, Davis, and folded, stuffed, and mailed out on June 5, 2006 by the UCD Bulk Mail Center.

The identification number mentioned in the letter is a unique 10-digit code that contains the respondent's census tract, street address, and unique identifying number (for those in apartment complexes). The code uses a private formula to disguise the linkage to location from any unauthorized observer: the first two digits are unique to each of the 35 census tracts in the target sample; digits 3 through 7 contained a scrambled version of the street address, wherein the address numbers are converted one to one (0=8, 1=4, 2=6, 3=0, 4=7, 5=3, 6=5, 7=2, 8=1, 9=9), and the order is scrambled (3^{rd} digit, 1^{st} digit, 4^{th} digit, 2^{nd} digit, 5^{th} digit), with 0s filling in the blanks, and the 0's are then converted to 8s, so that 1234 becomes 84607→68047. The reason for devising such a code was to preserve useful information (rather than being just a random identifying number) in case the correspondence between the code and the addresses was misplaced or disturbed. The web-based survey instrument required the respondent to enter the 10-digit code before proceeding with the survey (though any number, or string of characters, for that matter, could be entered in this location of the survey). The paper-based surveys had the 10-digit numbers recorded on them before being mailed out. However, the codes resulted in numerous entry errors on the part of the internet respondents – most of which could be identified and fixed, but not all. In retrospect, the benefit of the code probably did not exceed its cost.

Approximately two weeks after sending out the recruitment letter, a postcard reminder was supposed to be sent to everyone who had not completed the survey by the time the information had to be provided to Bulk Mail (which turned out to be everyone except the 320 individuals who completed the survey within the first couple of days). One week later, a second postcard reminder was to be sent out to those who had not yet completed the survey. Accordingly, we had 7,800 copies of the first postcard printed in advance, and 7,600 copies of the second. Due to an error in the mail room, however, the second postcard was labeled for mailing first, and we were notified when they ran out of postcards (having only 7,600 instead of the 7,680 needed for the first mailing). Once the mistake was identified, the 7,680 first postcards were labeled and mailed out. A week later, the 7,600 erroneously labeled second postcards were mailed (without removing those who had returned the survey, fearing that that could lead to further errors), leaving the last 80 names on the first-reminder mailing list without the second reminder postcard.

UNIVERSITY OF CALIFORNIA, DAVIS

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

Patricia L. Mokhtarian, Professor
Department of Civil and Environmental Engineering
and Associate Director, Institute of Transportation Studies
One Shields Avenue
Davis, California 95616-5924

PHONE: (530) 752-7062
FAX: (530) 752-7872 or -6572
plmokhtarian@ucdavis.edu
www.its.ucdavis.edu/telecom/

June 1, 2006

Dear Davis resident,

Nearly everyone does some shopping from time to time, whether for groceries, music, clothing, tools, or other items. The Institute of Transportation Studies at the University of California, Davis is conducting a study to learn more about how people shop - whether in stores, by catalog, or over the internet. By improving our knowledge about shopping trends, we can better predict future shopping patterns, and understand what it may mean for traffic patterns in cities.

We are hoping that you will be able to help, by sharing your own views and experiences with us. Your neighborhood is one of two in Davis that have been selected for this study, and your address was randomly selected within your neighborhood. Your participation is entirely voluntary and confidential: responses will never be linked to individuals.

Because we are only sending this invitation to a small sample, your response is extremely important. We are interested in your answer to each question *regardless of how much or how little you shop or what you shop for*.

The survey is being conducted on the internet. To complete the survey, please enter the address http://www.its.ucdavis.edu/research/survey.htm into your favorite web browser (such as Internet Explorer, Firefox, or AOL). To restrict the survey to our sample, you will also need the following ID number: 04-28842-001. If you would rather complete a paper version of the survey, please contact us as indicated below and we will be happy to mail you the survey with a postage-paid business reply envelope.

To help us randomly select the respondent to this survey, we are asking that the adult (16 years or older) in your household whose birthday is closest to June 1st complete the survey. If that person is unwilling to complete the survey, another adult in the household is welcome to do so. The survey should take about half an hour to complete (depending on connection speed), and we think you'll find it interesting and fun to do.

Everyone who completes the survey by ***June 21, 2006*** will be entered into a drawing for ***5 cash prizes of $100 each***. Because of the limited number of surveys that will be completed, your chances of winning will seldom be higher! If you are unable to fill out the questionnaire by June 21, 2006, please complete it as soon as possible.

Thank you in advance for your participation in this valuable study. If you are interested in our results, they will appear in the future on my web site at www.its.ucdavis.edu/telecom. If you have any questions, feel free to contact me at (530) 752-7062 or *plmokhtarian@ucdavis.edu*, or my assistant Mr. David Ory at (530) 754-7217 or *dtory@ucdavis.edu*.

Sincerely,

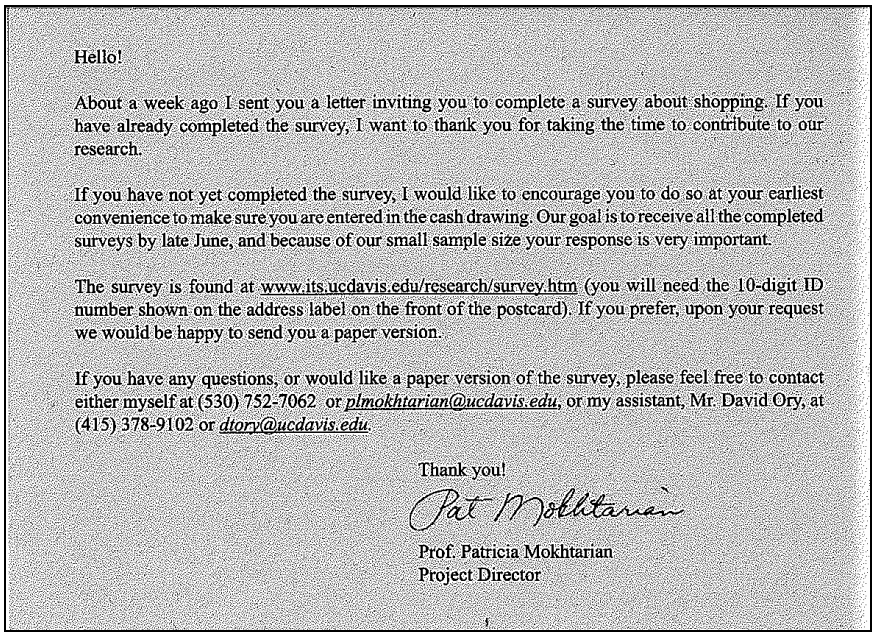
Patricia L. Mokhtarian

Professor Patricia L. Mokhtarian, Project Director

http://www.its.ucdavis.edu/research/survey.htm ID: 04-28842-001

Figure 4: Davis Recruitment Letter

The first postcard reminder was printed on mint green card stock and is included as Figure 5 below. It was mailed on June 20, 2006. The second postcard reminder, shown in Figure 6, was printed on bright yellow card stock and mailed on June 26, 2006.



Hello!

About a week ago I sent you a letter inviting you to complete a survey about shopping. If you have already completed the survey, I want to thank you for taking the time to contribute to our research.

If you have not yet completed the survey, I would like to encourage you to do so at your earliest convenience to make sure you are entered in the cash drawing. Our goal is to receive all the completed surveys by late June, and because of our small sample size your response is very important.

The survey is found at www.its.ucdavis.edu/research/survey.htm (you will need the 10-digit ID number shown on the address label on the front of the postcard). If you prefer, upon your request we would be happy to send you a paper version.

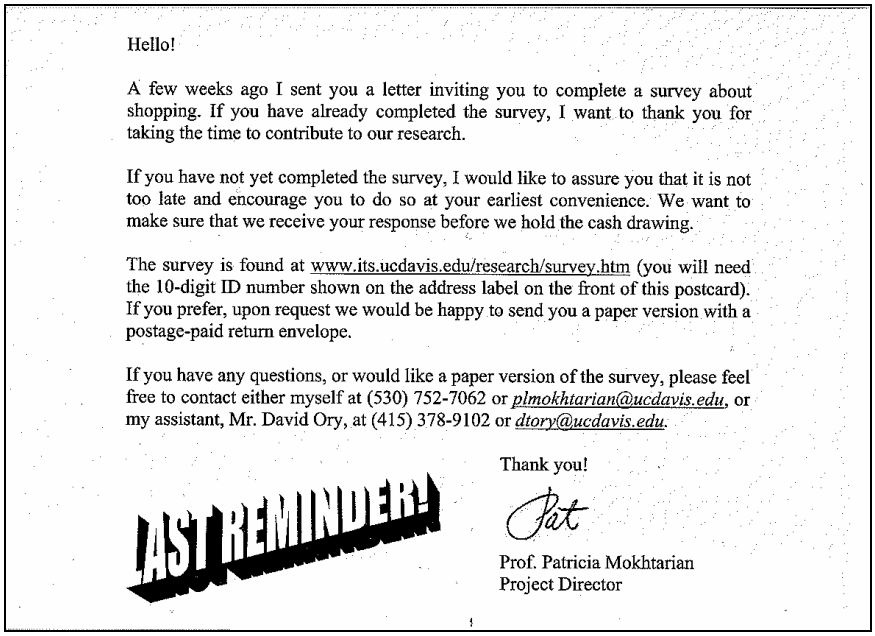
If you have any questions, or would like a paper version of the survey, please feel free to contact either myself at (530) 752-7062 or *plmokhtarian@ucdavis.edu*, or my assistant, Mr. David Ory, at (415) 378-9102 or *dtory@ucdavis.edu*.

Thank you!

Pat Mokhtarian

Prof. Patricia Mokhtarian
Project Director

Figure 5: First Postcard Reminder



Hello!

A few weeks ago I sent you a letter inviting you to complete a survey about shopping. If you have already completed the survey, I want to thank you for taking the time to contribute to our research.

If you have not yet completed the survey, I would like to assure you that it is not too late and encourage you to do so at your earliest convenience. We want to make sure that we receive your response before we hold the cash drawing.

The survey is found at www.its.ucdavis.edu/research/survey.htm (you will need the 10-digit ID number shown on the address label on the front of this postcard). If you prefer, upon request we would be happy to send you a paper version with a postage-paid return envelope.

If you have any questions, or would like a paper version of the survey, please feel free to contact either myself at (530) 752-7062 or *plmokhtarian@ucdavis.edu*, or my assistant, Mr. David Ory, at (415) 378-9102 or *dtory@ucdavis.edu*.

LAST REMINDER!

Thank you!

Pat

Prof. Patricia Mokhtarian
Project Director

Figure 6: Second Postcard Reminder

The web surveys were "active" from June 1 to September 14, collecting a total of 996 unique responses (complete and partial). Combined with 71 returned paper surveys (out of 80 mailed out), a total of 1,067 surveys were returned (note: these numbers and all those in tables below do not include one additional paper survey returned months later, but still in time to include in future analyses). To compute a response rate for the surveys, we first subtracted the number of "bad" addresses provided by the commercial vendor. These were computed by tallying (by city) the number of recruitment letters returned as undeliverable. A total of 1,426 letters were returned to the University, indicating that 17.8% of the addresses supplied by the vendor were out-of-date in terms of the name of the current resident (not surprisingly in view of the mobility of its heavily college-related population, the bad address rate was higher in Davis than in Santa Clara). Subtracting this number from 8,000 leaves 6,574 households who (presumably) received our correspondence. Of the 6,574, 1,068 completed the survey, resulting in a response rate of approximately 16 percent.

Table 10 summarizes the response rate by city. Interestingly, the response rate for Davis residents is considerably higher (at near 23%) than for Santa Clara residents (at near 9%). We attributed this difference to two factors. First, it is likely that the affection Davis residents have for the University of California campus in their city, which accounts for the lion's share of economic activity in the city, accounted for the majority of the boost in response rate. Also, by coincidence the survey corresponded with a debate in the city over the proposed construction of a Target Store (no so-called "big-box" retail stores are currently located in Davis). It is likely (anecdotally corroborated by conversations with several respondents and by several comments written on the survey) that those with strong opinions towards the Target store saw the survey as an opportunity to share their views about the issue.

Table 10: Response Rate by City

Quantity	Davis	Santa Clara	Unknown	Total
Mailed out	4,000 (100%)	4,000 (100%)	---	8,000 (100%)
Undeliverable mail	754 (18.9%)	674 (16.9%)	---	1,428 (17.9%)
Submitted (complete and partial) surveys	760	297	10**	1,067
Response rate*	23.4%	8.9%	---	16.2%

* Percent of delivered letters resulting in a submitted survey. **ID codes entered incorrectly so location could not be ascertained, but survey appeared to be legitimate.

The first web-based survey was submitted on June 9; the last survey was submitted on August 22, 2006. Figure 7 shows the distribution of completed surveys over time. Note that the chart does not include partial responses, as they are not time-stamped by the website. Paper surveys entered into the web-based survey, by us, are also omitted from the distribution.

More than 13% of all completed responses were collected on the first day and over half by the end of the first week. The impact of both postcard reminders is evident in the chart: more than 40% of all responses were received after the first reminder was mailed, with almost 20% of the total occurring after the second reminder. However, the timing of the first reminder corresponded with the stated deadline for entering the raffle for the cash prize, and thus that "jump" in the response rate could be the result of both effects. The data collection was essentially complete after one month.

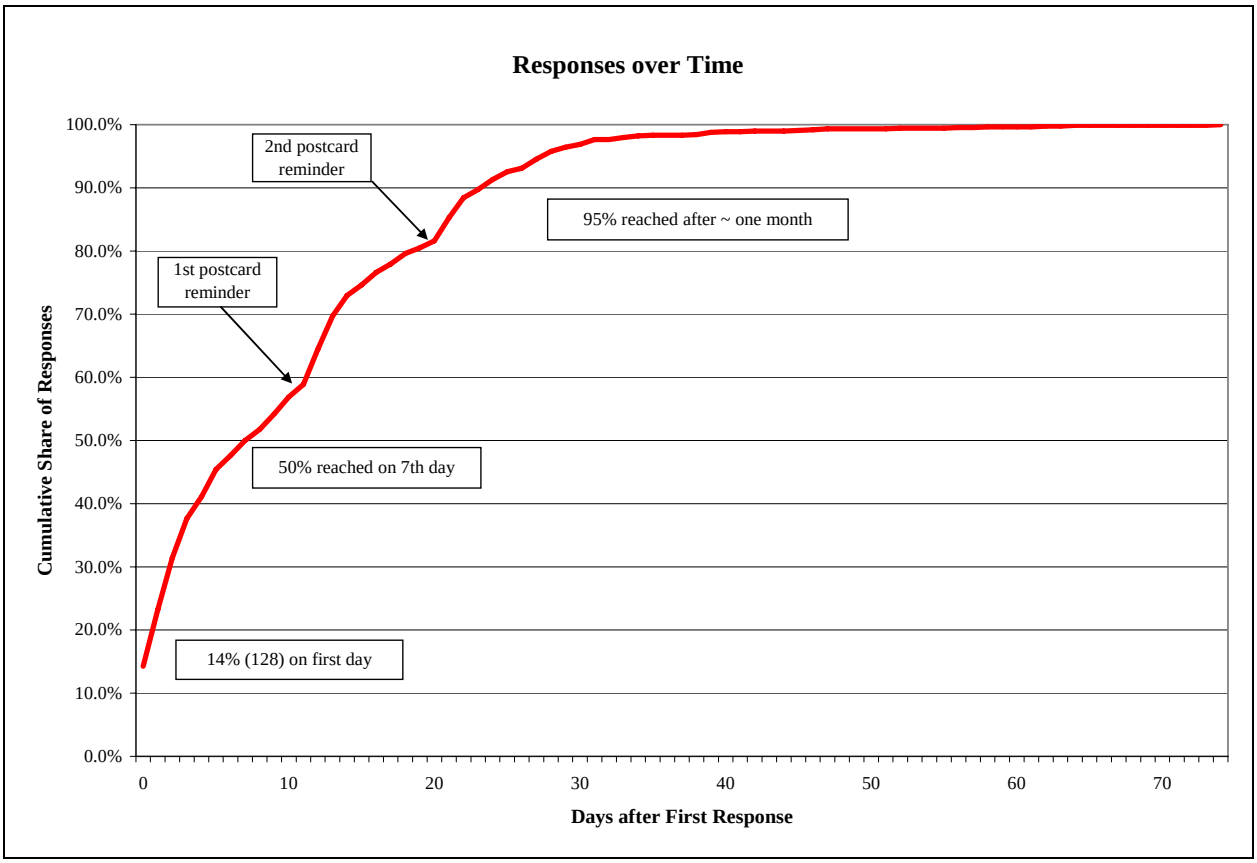


Figure 7: Completed Surveys over Time

7. INITIAL DATA CLEANING

The initial data cleaning activities focused on three aspects of the survey: the attitudinal statements in Part A and D, and the sociodemographic data in Part G. As the attitudinal statements allow for the most interesting analyses of the data and the sociodemographics allow for behaviors to be related to general populations, these portions of the survey are viewed as the most crucial.

Part A contains 42 attitudinal statements to which respondents can either agree or disagree on a five-point Likert-type scale. The distribution of completed responses is shown in Table 11. A natural cut-off was 3 or fewer missing data items, after which the distribution thins out dramatically. As such, those missing more than 3 responses were excluded from the working sample. Those missing three or fewer responses had their missing data imputed with geographically-segmented means. Using census tracts, the sample was segmented into six imputation groups (see Table 12). Mean values on a given statement were then computed for each of these groups, rounded to the nearest valid answer (the integers between 1 and 5, inclusive) and imputed accordingly.

As can be deduced from Table 11, only 108 out of 1,033 x 42 = 43,386 responses (0.25%) were imputed in this way. No single question in Part A had more than 8 (0.77% of 1,033) responses imputed.

Table 11: Part A Missing Data Summary

Missing	Valid	Frequency	Percent	Cumulative Percent
0	42	946	88.7%	88.7%
1	41	72	6.7%	95.4%
2	40	9	0.8%	96.3%
3	39	6	0.6%	96.8%
5	37	2	0.2%	97.0%
7	35	2	0.2%	97.2%
8	34	1	0.1%	97.3%
9	33	1	0.1%	97.4%
14	28	4	0.4%	97.8%
28	14	4	0.4%	98.1%
31	11	4	0.4%	98.5%
42	0	16	1.5%	100.0%
Total		1,067	100.0%	

Table 12: Geographic Imputation Groups

Imputation Group	N	Census Tracts
1	281	105.XX
2	335	106.XX
3	144	107.XX
4	48	5049.XX
5	155	5050.XX to 5053.XX
6	94	5054.XX to 5060.XX

Key: “XX” represents any number, i.e. 105.XX = 105.01, 105.02, 105.03, etc.

A similar selection and imputation process was performed on the Part D data. Recall from the previous section that Part D is segmented by product type and shopping mode. Part D, Section 1 contains the questions regarding store purchases (for both books (search) and clothing (experience)); Section 2 contains questions regarding either internet or catalog purchases – again for both item types. The distribution of these two sections are shown in Table 13 (Section 1) and Table 14 (Section 2). A cut-off of

three missing was again implemented. Those missing more than three responses in either section were excluded from the working sample.

The imputed means were specific to geography (using the groups of Table 12), item type (book or clothing), and shopping mode (for Section 2 only – catalog or internet). Only 77 of 991x28 = 27,748 responses (0.28%) in Section 1 (store) were imputed in this way, with no single question in Section 1 having more than 6 responses (0.61% of 991) imputed. Only 91 of 981x28 = 27,468 responses (0.33%) in Section 2 (internet or catalog) were imputed in this way, with no single question in Section 2 having more than 8 responses (0.82% of 981) imputed.

Table 13: Part D, Section 1 Missing Data Summary

Missing	Valid	Frequency	Percent	Cumulative Percent
0	28	925	86.7%	86.7%
1	27	57	5.3%	92.0%
2	26	7	0.7%	92.7%
3	25	2	0.2%	92.9%
6	22	2	0.2%	93.1%
7	21	1	0.1%	93.2%
14	14	6	0.6%	93.7%
16	12	1	0.1%	93.8%
28	0	66	6.2%	100.0%
Total		1,067	100.0%	

Table 14: Part D, Section 2 Missing Data Summary

Missing	Valid	Frequency	Percent	Cumulative Percent
0	28	906	84.9%	84.9%
1	27	61	5.7%	90.6%
2	26	12	1.1%	91.8%
3	25	2	0.2%	91.9%
5	23	1	0.1%	92.0%
12	16	1	0.1%	92.1%
14	14	4	0.4%	92.5%
15	13	1	0.1%	92.6%
28	0	79	7.4%	100.0%
Total		1,067	100.0%	

Though numerous demographic measures were captured by the survey, the most important were felt to be: gender, number of vehicles, number of workers, household size, employment status, and income. Taking each of these six questions as an individual item, the distribution of missing items is shown in Table 15. The seemingly magic number of three was again used as a cut-off: those missing more than three of these six items were excluded from the working sample. No imputation was done for the missing values.

Table 15: Sociodemographic Missing Data Summary

Missing Items	Valid Items	Percent	Cumulative Percent
0	856	80.2%	80.2%
1	118	11.1%	91.3%
2	23	2.2%	93.4%
3	8	0.7%	94.2%
4	1	0.1%	94.3%
5	4	0.4%	94.7%
6	57	5.3%	100.0%
Total	1,067	100.0%	

Using the cut-off criteria for Part A, Part D, and the sociodemographics, our original sample of 1,067 responses was reduced to 966 (91% of the original total) via a sequential process illustrated in Figure 8. The 966-observation file will become the “working file” on which all subsequent work will be based. Table 16 through Table 19 summarize the missing and/or imputed data distribution for the final working file.

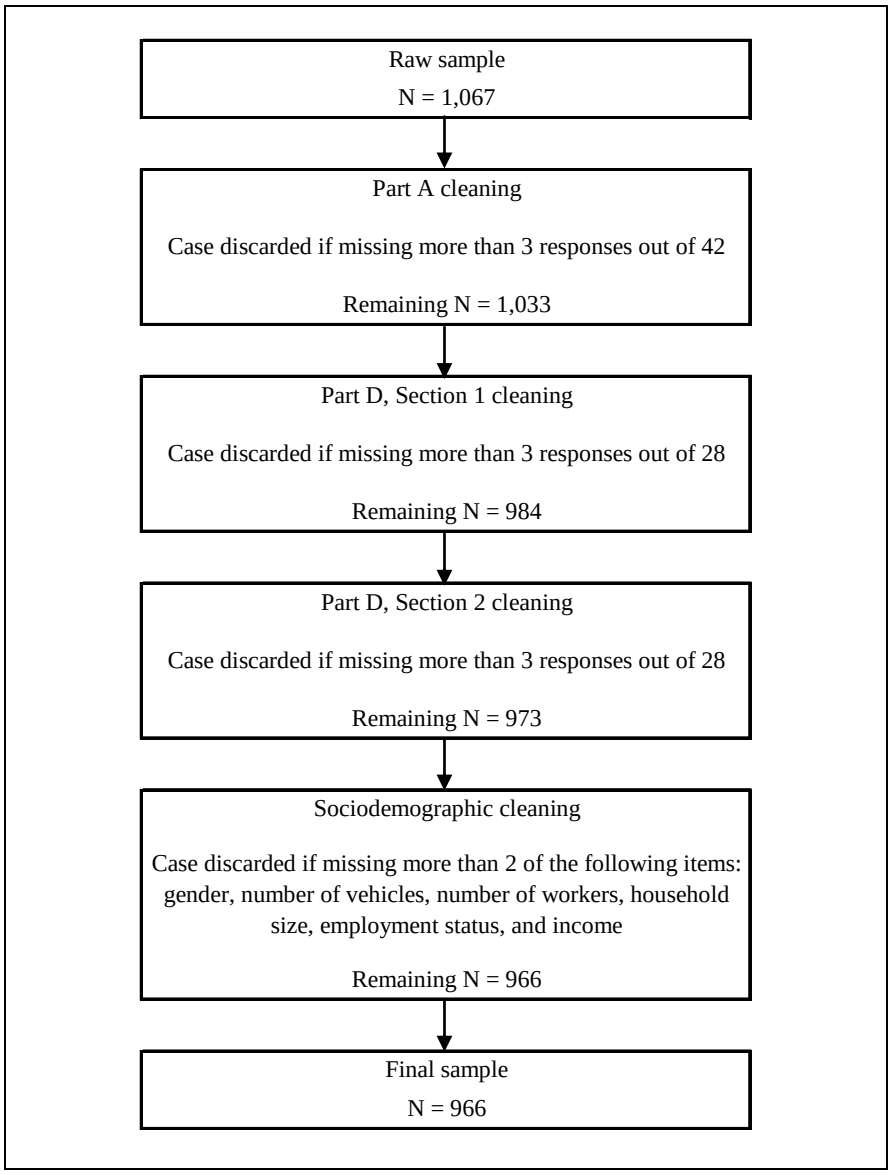


Figure 8: Data Cleaning Process

Table 16: Part A Missing (Imputed) Data Summary for Final Sample

Missing	Valid	Frequency	Percent	Cumulative Percent
0	42	889	92.0%	92.0%
1	41	65	6.7%	98.7%
2	40	8	0.8%	99.6%
3	39	4	0.4%	100.0%
Total		966	100.0%	

Table 17: Part D, Section 1 Missing (Imputed) Data Summary for Final Sample

Missing	Valid	Frequency	Percent	Cumulative Percent
0	28	904	93.6%	93.6%
1	27	55	5.7%	99.3%
2	26	6	0.6%	99.9%
3	25	1	0.1%	100.0%
Total		966	100.0%	

Table 18: Part D, Section 2 Missing Data Summary for Final Sample

Missing	Valid	Frequency	Percent	Cumulative Percent
0	28	897	92.9%	92.9%
1	27	58	6.0%	98.9%
2	26	9	0.9%	99.8%
3	25	2	0.2%	100.0%
Total		966	100.0%	

Table 19: Sociodemographic Missing Data Summary for Final Sample

Missing Items	Valid Items	Percent	Cumulative Percent
0	828	85.7%	85.7%
1	116	12.0%	97.7%
2	22	2.3%	100.0%
Total	966	100.0%	

Table 20: Target Status of Working File Cases

Measure	Santa Clara		Davis	
	Traditional	Suburban	Traditional	Suburban
On-target addresses	41	53	109	237
Off-target addresses	177		346	

8. SUMMARY AND NEXT STEPS

The survey design and data collection effort can be considered reasonably successful, producing a relatively clean dataset with enough cases to permit the application of numerous statistical methods. The dataset is extraordinarily rich with attitudinal variables, as well as having a number of different behavioral indicators and the conventional sociodemographic traits. In fact, these data offer the most comprehensive set of shopping-related variables that we have seen empirically measured by a single study. Accordingly, we believe they will continue to provide useful insights for some time to come, especially with respect to the role the internet is playing in the shopping behavior of ordinary Americans.

Initial analysis plans call for factor-analyzing the Part A and Part D attitudinal statements, and clustering cases on factor scores to identify market segments having similar attitudinal profiles. It is also of interest to examine attitudes differ by mode as well as product type (book/CD/DVD/videotape, a search good, versus clothing/shoes, an experience good). Eventually, discrete choice models will be developed using a variety of dependent variable formulations (actual and intended choices for a single purchase, as well as frequencies and shares for multiple purchases), and beyond that, applications of more sophisticated methodologies such as latent class and structural equations models await.

REFERENCES

Babbie, Earl (1998) *The Practice of Social Research*, 8th edition. Belmont, CA: Wadsworth Publishing.

Baumgartner, Hans and Jan-Benedict E. M. Steenkamp (2001) Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research* **38**(May) 143-156.

Brownstone, David (1998) Multiple imputation methodology for missing data, non-random response, and panel attrition. Chapter 18 in Tommy Garling, Thomas Laitila, and Kerstin Westin, eds. *Theoretical Foundations of Travel Choice Modeling*. Amsterdam: Elsevier, 421-449.

Brynjolfsson, Erik and Michael D. Smith (2001) The great equalizer? Consumer choice behavior at Internet shopbots. Working paper, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, April. Available at ebusiness.mit.edu/papers/tge.

Burke, Raymond R. (1998) Real shopping in a virtual store. In *Sense and Respond: Capturing Value in the Network Era*. Eds. Stephen P. Bradley and Richard L. Nolan. Boston: Harvard Business School.

Cairns, Sally (1996) Delivering alternatives: Successes and failures of home delivery services for food shopping. *Transport Policy* **3(4)**, 155-176.

Cao, Xinyu and Patricia L. Mokhtarian (2005) *The Intended and Actual Adoption of Online Purchasing: A Brief Review of Recent Literature*. Research Report UCD-ITS-RR-05-23, Institute of Transportation Studies, University of California, Davis, May.
Available at http://repositories.cdlib.org/itsdavis/UCD-ITS-RR-05-07/.

Ellard, John H. and T. B. Rogers (1993) Teaching questionnaire construction effectively: The Ten Commandments of question writing. *Contemporary Social Psychology* **17(1)** (March), 17-20.

Fabrigar, Leandre R., Duane T. Wegener, Robert C. MacCallum, and Erin J. Strahan (1999) Evaluating the use of exploratory factor analysis in pyschological research. *Psychological Methods* **4(3),** 272-299.

Farag, Sendy, Kevin J. Krizek, and Martin Dijst (2006) E-shopping and its relationship with in-store shopping: Empirical evidence from the Netherlands and the USA. *Transport Reviews* **26**, 43-61.

Ferrell, Christopher E. (2005) Home-based teleshopping and shopping travel: Where do people find the time? *Transportation Research Record* **1926**, 212-223.

Gould, Jane and Thomas F. Golob (1997) Shopping without travel or travel without shopping? An investigation of electronic home shopping. *Transport Reviews* **17(4)**, 355-376.

Gould, J., T. Golob, and P. Barwise (1998) Why do people drive to shop? Future travel and telecommunication trade-offs. Presentation at the Annual Meeting of the Transportation Research Board, Washington, DC, January.

Handy, S. and T. Yantis (1997) *The Impacts of Telecommunications Technologies on Nonwork Travel Behavior.* Research Report SWUTC/97/721927-1F, Southwest Region University Transportation Center, Center for Transportation Research, The University of Texas at Austin, January.

Krizek, Kevin J., Yi Li, and Susan L. Handy (2005) Spatial attributes and patterns of use in household-related ICT activity. Paper no. 05-1933 on the conference CD of the 84th Annual Meeting of the Transportation Research Board, Washington DC, January.

Koppelman, Frank, Ilan Salomon, and Kimon Proussaloglou (1991) Teleshopping or store shopping? A choice model for forecasting the use of new telecommunications-based services. *Environment and Planning B: Planning and Design* **18**, 473-489.

Peterson, Robert A., Sridhar Balasubramanian, and Bart J. Bronnenberg (1997) Exploring the implications of the internet for consumer marketing. *J. of Academy of Marketing Science* **25 (4)**, 329-346.

Ren, and Mei-Po Kwan (2005) The impact of e-shopping on individual temporal constraints and out-of-home activities. Paper 05-1647 on the CD-ROM of the 84th Annual Meeting of the Transportation Research Board, January, Washington DC.

Salomon, Ilan and Frank S. Koppelman (1988) A framework for studying teleshopping versus store shopping. *Transportation Research A* **22(4)**, 247-255.

Tacken, M. (1990) Effects of teleshopping on the use of time and space. *Transp. Res. Rec.* **1285**, 89-91.

Tauber, E. (1972) Why do people shop? *Journal of Marketing* **36**, 46-49.