

CHAPTER 1

Survey Interviews and New Communication Technologies

Michael F. Schober

New School for Social Research, New York, New York

Frederick G. Conrad

University of Michigan, Ann Arbor, Michigan

1.1 INTRODUCTION

How people communicate with each other today—at least those with access to new technologies—is enormously different from how they communicated with each other even ten years ago. Internet-based cell phones, desktop videoconferencing, mobile instant messaging, blogs, podcasts, and other current modes of communication would have been unimaginable, or the stuff of science fiction, for prior generations. And the pace of change only seems to be increasing, with new generations embracing and creating newer forms of communication.

The pace of this change poses a substantial challenge to one of the institutions on which modern societies rest: the collection of data in survey interviews. It is through individual survey interviews that societies track their employment, health, crime, demographic shifts, and citizens' opinions about controversies of the day, among many other social phenomena. The moment in which a survey respondent provides an answer to a survey question is thus far more consequential than it might at first seem, and what takes place at that moment is in a state of major transition. Changes in communication technology have already led to questions about the accuracy and generalizability of large-scale survey data. For example, as cell phone adoption increases,

it is unclear whether the data collected via traditional landlines are still as reflective of the entire population, and it is unclear whether social researchers will continue to be able to sample populations in ways that reliably generalize if they cannot interview respondents' whose only phone is a mobile phone. Yet cell phones also introduce new problems, such as the effects of answering questions in public, or while driving (see Fuchs, Chapter 4 in this volume).

The urgency of understanding and anticipating how new technologies will affect respondents' willingness to participate and their ability to provide accurate answers is acute. The barrage of requests that people receive to provide information has increased precipitously, as market researchers and opinion pollsters of varying degrees of legitimacy vie with census bureaus and social scientists to gain access to respondents. For these and other reasons (e.g., see Groves and Couper, 1998; Groves et al., 2002), people are becoming increasingly reticent to participate in survey interviews, and a greater percentage are opting out altogether. The implications become disturbing if we consider the point made by Ken Prewitt (2006), political scientist and former director of the U.S. Census Bureau: if governments become unable to gather essential information about citizens through interviews, in which the citizens have consented to participate, they will start to depend on surveillance data that citizens have *not* consented to provide. It is already possible to track a great deal about people's behavior (though probably not their opinions) from credit card usage, travel reservations, police and hospital records, and so on, and about their tastes and preferences from purchase profiles. Just how far societies develop into regimes of surveillance will depend, in part, on the degree to which survey interviews continue to be the major source of societal data.

Contending with the arrival of new technologies that affect communication is not new to survey research. Historically, survey researchers have, in fact, been key innovators in creating new technologies; for example, Hollerith invented the all-important punch card at the Census Bureau in 1890. In a more general sense, the great success of telephone surveys in the 1970s, especially after the incorporation of computer support in the 1980s, has served as a model for other mass scale telephone transactions. For example, telemarketing (sidestepping for the moment an evaluation of its social impact) and reservation centers owe at least an indirect debt to the pioneers of telephone surveys.

Yet more recently, survey practitioners have tended to follow rather than lead in devising and adopting new technologies. Quite some time after desktop computers with graphical user interfaces had become ubiquitous worldwide, survey practitioners continued to report on their plans for the transition from DOS to Windows. The same is true for the introduction of laptop computers into field data collection. Cell phones are another example: many years after they have become the primary phones in many homes, survey researchers still grapple with whether and how to incorporate them into telephone surveys. While speech and handwriting recognition tools are becoming more commonplace (e.g., speech interfaces to reservation systems and in-car GPSs; handwriting recognition in hand-held computers), they still occupy only a small niche in survey methodologists' suite of tools. Although the Internet is increasingly used for web surveys, its adoption lags far behind that for other activities like e-commerce,

electronic tax filing, and so on.¹ And survey researchers have hardly considered the potential use and impact of mobile instant messaging, desktop videoconferencing, animated agent (avatar) systems, and other technologies whose role in daily communication is increasing rapidly.

There are legitimate reasons for conservative adoption of new technologies by survey practitioners. Given the extensive infrastructure costs, survey practitioners may seek to avoid moving to less proven approaches from technologies that are working reliably and that produce data that fulfill clients' needs. Survey practitioners are rightly concerned about the quality of the statistical estimates produced by surveys. If the incomplete penetration of a technology or the lack of complete sample frames leads to unknown error, adopting the technology could be risky.

Despite these legitimate concerns, we have reached the point of needing to be more proactive in trying to understand how current communication technologies, and those on the horizon, are used by humans and how their use in surveys would affect data quality. Because preferred ways of communicating are changing on a large scale, adhering to tried and true methods for survey interviewing is likely to lead to loss of participants and data quality; some potential respondents may already be unwilling to participate with today's standard technology (e.g., landline telephones). For example, people who spend much of the day multitasking on their smart phones may be more likely to interact with an interviewer via instant messaging than by talking. The point is that we can no longer afford *not* to consider the communicative properties of new technologies. This will require developing theories that allow us to exploit the characteristics of new media, rather than simply emulate familiar practice with new tools.

We also will need to move beyond assuming that design principles from one medium are the right principles for another. So, for example, rather than simply applying the design principles developed for paper questionnaires in web-based instruments, because these are well known and vetted (e.g., see Dillman, 2000), we propose that we should capitalize on the unique properties of web interaction. For example, one can imagine web questionnaire interfaces that adapt or tailor themselves to respondents' behavior, that diagnose respondents' need for clarification, and that allow respondents to request clarification (Conrad, Schober, and Coiner, 2007; Schober, Conrad, Ehlen, Lind, and Coiner, 2003). Other visual and interactive aspects of web surveys are also exploitable, with systems that detect respondents' lack of effort, that instantly check some answers against electronic records, and other sorts of features on the horizon.

Because what is cutting edge today will not be so for long, what is even more essential is developing principles and theories that remain useful and applicable irrespective of what technology is on the horizon at any point in time. This will require far more extensive connection and conversation between communication technologists and survey researchers than currently occurs.

¹Of course, the barriers to widespread deployment of web surveys have at least as much to do with incomplete access to the Internet within the population as they do with the technology itself (see Chapters 3, 4, and 12 in this volume).

This chapter, and this volume, represent an attempt to foster this conversation. For survey researchers, the conversation requires learning a bit more not only about features of technologies, but also the principles and taxonomies that guide the development of those technologies. For communication technologists, the conversation requires learning a bit more about the survey enterprise, the nature of interviewing, and what survey designers would need to be assured of before being willing to adopt a new technology. Here we will present brief introductions to what each community should know in order to engage in the conversation. We also raise a set of questions that need to be addressed when considering whether to adopt a new communication technology. Some of these questions are relatively straightforward to answer, but answering others will require knowing the results of research that hasn't yet been carried out.

We should also be clear about what we are *not* doing. Although there are various important technological developments that peripherally affect the survey interview, this chapter and volume will focus on communication technologies in the survey interview itself. Thus, we will not focus on new technologies for constructing samples; non-interview methods for collecting data (e.g., passive measurement devices that eliminate respondents' self-report, like devices that detect cigarette smoke for a survey on smoking); or questionnaire construction, question development, and survey translation (into different languages), which happen before the interview. All these issues deserve serious attention, but, again, the focus here is on the interview itself.

1.2 SURVEY INTERVIEWS: THE STATE OF THE ART TODAY

A primary distinction in how surveys are administered today is between interviewer- and self-administration (see Couper, Chapter 3 in this volume). As we will argue, for considering the next generation of survey data collection, this distinction is becoming hazier.

The most frequent medium in which interviewer-administered surveys occur today is via landline telephone calls; face-to-face interviews still occur, but they are not the norm as they were fifty years ago. Typically, interviewers call from call centers where they have been carefully trained, and their performance is often monitored, either live or through random recording. Interviewers are most often trained to implement some version of standardized interviewing (see Fowler and Mangione, 1990), in which interviewers read scripted questions exactly as worded and make sure not to deviate from their script. In the strictest forms of standardized interviewing, interviewers leave the interpretation of questions entirely up to respondents; they do not define words in questions even if respondents ask for clarification, because departing from the script for some respondents but not others could lead to different interviewers affecting the answers differently. These days interviewers most often sit in front of computer screens (computer-assisted telephone interviewing, or CATI), so that a program prompts them with the questions they are to deliver, fills in specifics like the respondent's name, collects the answers they enter during the interview, and can do some error-checking and skip to relevant next questions based on answers to prior questions.

The energy and expenditure that go into maintaining call centers, and into selecting, training, and monitoring interviewers, are enormous. In fact, interviewer costs are often the greatest expense for carrying out a survey (although this is not to suggest that interviewers, many of whom work part time, are highly paid). Finding interviewers who sound natural on the phone, who are good at enticing respondents to participate, who manage to follow the script while maintaining rapport, and who are careful and meticulous enough to treat the data with care even at the end of a long shift is not a trivial task. And, while inferable (from audio) characteristics of interviewers like their age, gender, dialect, or ethnicity can affect respondents' willingness to participate and the answers that respondents give (Ehrlich and Riesman, 1961; Hatchett and Schuman, 1975; Kane and Macaulay, 1993; Schuman and Converse, 1971), it is not usually within the control of survey centers what the characteristics of their interviewers turn out to be.

The other major form of survey data collection these days is self-administration, via paper and pencil in mail-in surveys like the U.S. Census, but more and more via text-based web surveys. From survey designers' point of view, there are a number of advantages and disadvantages to self-administration. Advantages include both the cost savings of not needing interviewers and the fact that, if there are no interviewers, then different interviewers can't have differential effects on the data. Respondents can also choose to respond when it is convenient for them, and without time pressure that might distort their responses. Disadvantages include the greater difficulty of persuading respondents to participate, as well as the impossibility of establishing rapport with respondents, which might convince them to consider response options and complete the survey carefully.

Newer forms of administration begin to blur the lines between interviewer- and self-administration. Consider Interactive Voice Response surveys, where respondents listen to recorded voices on their telephone and speak or keypad their answers (see Bloom, Chapter 6 in this volume). On the one hand, from the survey designers' perspective this is clearly self-administration, in that no live interviewers are involved and the interface is perfectly standardized. From respondents' perspective, however, the interaction with the survey instrument unfolds over time in ways that are more similar to interviewer-administration, with spoken questions that sound as they would if a live interviewer were reading them, and with a turn-taking structure akin to survey discourse. Or consider Audio Computer-Assisted Self-Interviewing (A-CASI), in which a face-to-face interviewer turns the laptop over to the respondent for sensitive or threatening portions of interviews; the respondent listens to audiorecorded questions over headphones and types answers into the laptop. This provides privacy, so that other household members and even the interviewer don't know what the answers are, not to mention the domain of questioning; it can lead respondents to answer more honestly (e.g., Tourangeau and Smith 1996). Again, survey designers classify this as self-administration, but the interactive dialogue from the respondent's perspective bears notable similarity to interviewer-administration.

These more ambiguous cases only hint at the level of complexity that is on its way. Self-administered interviewing is likely to continue to add more and more features of human interviewing, turning the distinction between self- and interviewer-administration into a continuum or even, eventually, making the distinction obsolete.

Interviews with human interviewers that are mediated by videoconferencing certainly include interviewers, but they are removed from full physical copresence. Textual web interfaces and speech interfaces that allow respondents to request clarification by clicking or asking are under development (e.g., Conrad, Schober, and Coiner, 2007; Conrad et al., 2006; Ehlen et al., 2007); these could incorporate a level of interactivity that goes beyond what live interviewers trained in the strictest form of standardization are currently able to deliver. Interviews by talking head animated agents, akin to the University of Memphis' functioning AutoTutor tutoring system, have demonstrated effects of the ethnic appearance of the animated tutor (see Person, D'Mello, and Olney, Chapter 10 in this volume; and Cassell and Miller, Chapter 8 in this volume) that suggest the interaction may include more aspects of interviewer-administration than self-administration. Self-administration seems to lead to more candid responses to sensitive questions, but it is unclear at this point whether respondents will consider an interviewing system that includes an animated interviewing agent to be self- or interviewer-administered.

A next generation of interviewing systems is likely to make use of "paradata" (process measures) from respondents during the interaction, as a way of diagnosing when help might be needed so that the system can intervene appropriately to collect high quality data (see Couper, Chapter 3 in this volume; and Person, D'Mello, and Olney, Chapter 10 in this volume). One could imagine making use of respondents' typing errors, changed answers, response times, speech disfluencies, or facial expressions to assess their confidence in their answers or their likelihood of having misunderstood. Designers of interviewing systems will have to make decisions about when and how an interviewing agent should respond to this kind of user (respondent) behavior. Should the system respond during or after answers? What kind of audio or visual evidence should a computerized "interviewer" display? Should the interviewing agent say "okay" or "uh-huh" like human interviewers do? Should the interviewing agent smile or nod in response to an answer? When should the interviewing agent look at the respondent?

Considerations like these make it clear that survey researchers don't yet have the full set of computational and theoretical tools that it will take to make informed choices about the next generation of interviewing systems. And designers of other sorts of systems don't understand the unique constraints and requirements of survey interviewing. In the sections that follow, we present brief (and nonexhaustive) descriptions of what survey researchers need to know about what is going on in communication technology research and development and what communication technologists need to know about surveys.

1.3 WHAT SURVEY RESEARCHERS NEED TO KNOW ABOUT COMMUNICATION TECHNOLOGIES

People interact differently, in principled and characterizable ways, in different modes or media of communication. Scholars in a wide array of disciplines, including subareas of psychology, computer science, linguistics, communication, media studies,

human-computer interaction, management studies, information science, and sociology (among others), have proposed distinctions, observed behavior, and carried out experiments that are potentially useful for survey researchers who want to consider adopting a new communication technology.

The existing empirical base has, thus far, rarely come from studying survey interaction, but rather from examining people's behavior on a variety of real-world and laboratory-created tasks. These can range from studies of group decision-making in mediated business communication, to comparing people's instructions and references about maps or abstract shapes when they are videomediated versus interacting face to face or via instant messaging, to careful analyses of people's interactions with working animated agents in intelligent tutoring systems.

The extent to which existing work is relevant for survey interaction thus depends on how similar and different those tasks are from survey interviews. (This is why clearer characterization of today's survey interaction is extremely important as we move forward; see Schaeffer and Maynard, Chapter 2 in this volume, and papers in Maynard, et al., 2002). It also will depend on how convinced survey designers are about the generalizability of work that comes from case studies and samples of participants smaller than the large samples that surveys rest upon. Given that there are almost no large-sample studies available (carrying them out would be prohibitively expensive, and not clearly any more generalizable), survey designers will need a bit of interdisciplinary flexibility in order to extrapolate from the existing work.

Survey designers should know that communications theorists do not have a single agreed-upon taxonomy of the features that are relevant in understanding a communicative situation, although a number of important distinctions have been proposed, including:

- *Communicative Goals*. Interactive goals can be oriented toward completing the task (*task-oriented* communication) or toward the relationship with the conversational partner (*socioemotional* or *affective* communicative goals), or both.
- *Initiative*. Who "drives" the interaction?

Task versus Conversational Initiative. Different parties in an interaction may be in charge of the joint task (e.g., interviewers are responsible for collecting the data) and in charge of what is happening in the conversation at the moment (a respondent may take the conversational initiative by requesting that the question be re-read).

Single-Agent versus Mixed Initiative. In some interactions the task or conversational initiative comes from one party only (e.g., in standardized survey interviews the interviewer owns the task initiative), but in other kinds of interactions the initiative can pass back and forth (e.g., in potential interactive interviewing systems of the future).

Stakes. Some interactions are high stakes, where getting the details right matters (e.g., air-traffic control communication, negotiating child visitation rights), and others have lower stakes (e.g., small talk at a party, and

maybe, from survey respondents' perspectives, their answers during interviews).

Synchrony. In *synchronous* interactions, there is virtually no lag between conversational turns (e.g., on the telephone), but in *asynchronous* interactions there can be an indefinite lag (e.g., postings to blogs, voice mail messages, e-mail).

We see a number of areas of research on interaction in communication technologies that are potentially important for survey designers to know more about. This isn't the place for a detailed or exhaustive review, but here is some of the existing work, along with references for where to look further.

1.3.1 Comprehension in Different Media

People in conversation make sure they understand each other by confirming ("uh-huh," "okay," nodding) and questioning ("huh?", "what was that?", looking confused) during and after each other's utterances (see Clark, 1992, 1996; Schober and Brennan, 2003, among many others). The form of this "grounding" differs in different communication media, which *afford* different costs and constraints (see Clark and Brennan, 1991, and Whittaker, 2003, for useful and reader-friendly overviews). For example, whether participants talk, type, or click; whether they can see each other's faces or hear each other's voices, and whether the communication leaves a reviewable record are all affordances of a medium and can all affect how people ground their understanding (e.g., clicking is easier than typing; remembering is harder than being able to review). This works out differently for different kinds of communicative tasks. The kinds of findings in this area include evidence that over textual chat people coordinate more efficiently when using an explicit turn marker (like "o" for "over") than when such a marker is unavailable (Hancock and Dunham, 2001), and evidence that task efficiency can be relatively unaffected by availability of facial information (e.g., as afforded by video) when people are directing each other on maps (Anderson et al., 1997).

Beyond the verbal aspects of grounding understanding, nonverbal communicative displays are an important part of communication. Research and theorizing have focused on two main kinds of displays, which are differently available in different media. *Paralinguistic* aspects of communication are tied to the linguistic aspects; they include speakers' tone of voice, intonational contours, their fluency and disfluency, and pausing (for reviews and discussion, see Clark, 1996; Levelt, 1989). *Visual* displays are less directly connected with the language, although they are deeply intertwined; these include gesture, gaze, nods, and facial expressions (see Goodwin, 1991; Krauss et al., 1996; McNeil, 1992, among many others). Paralinguistic and visual displays can affect turn taking, referring, attention, and interpersonal impressions; this has mostly been demonstrated in comparisons of face-to-face and telephone conversations, but a growing literature is demonstrating how access to and lack of access to visual and audio cues can affect the nature of interaction in newer media (see Whittaker, 2003).

These cues are all potential paradata for survey researchers to make use of. Little is yet known about the reliability and utility of facial expressions and paralinguistic

displays in survey interviews, and there have been few comparisons across modes, but some work suggests that paralinguistic cues can be valid indicators of need for clarification. So, for example, survey respondents who pause longer and are more disfluent (*umming* more) before answering are more likely to be presenting answers that are inaccurate (Schober and Bloom, 2004); survey respondents who avert their gaze longer while answering questions in face-to-face interviews are more likely to be giving unreliable answers (Conrad, Schober, and Dijkstra, 2008).

1.3.2 Least Effort Strategies and Satisficing

As has been argued at least since Herbert Simon's classic work (1957), people tend to satisfy their problem-solving goals, including their interactional goals, with the least effort possible given their processing constraints. This phenomenon has been called "satisficing"—doing what it takes to get something done well enough, rather than working harder to maximize the outcome. The basic idea is that humans lack the cognitive resources (particularly working memory capacity) to consider all relevant information in making a decision and so make decisions that are suboptimal but often good enough. While the notion was developed to account for organizational decision making, satisficing and similar mental shortcuts have been observed across a vast range of human behavior.

In human-computer interaction, interface design has been shaped by the recognition that users will not engage in much of what they consider to be unnecessary action. For example, it has been shown (Gray and Fu, 2005) that when information that is needed to complete a task is displayed on the screen, people will avoid even a small eye movement, let alone physical interaction with the system, in order to obtain this information by *recalling* it—even though what they recall is less accurate than what is on the screen. A version of the satisficing idea has been applied to survey interviews (e.g., see Krosnick, 1991): a respondent can conceive of his/her goal as finishing the interview or questionnaire as quickly as possible rather than providing the most accurate data. Respondents might, for example, select the first adequate response option even though it isn't the best one, or provide similar answers to all questions in a grid (nondifferentiation), or respond before they have thought about an issue for very long. This obviously can lead respondents to provide suboptimal answers and is not what survey designers hope respondents will do.

Satisficing strategies have been shown to differ depending on what information is afforded by different communication media. For example, survey respondents are differentially likely to endorse earlier and later response options in a list when they are reading (earlier) rather than listening (later) (Krosnick and Alwin, 1986). Certainly whether questions are delivered as text or speech and whether the respondent controls the onset and timing of the question delivery will become only more important as new interviewing technologies are developed.

This suggests that, when survey researchers consider whether to adopt a new communication technology, they would do well to carry out the kind of task analysis advocated by Simon, thinking through how and when respondents can satisfice in

that medium. Media that promote additional satisficing are likely to be at odds with the goals of survey researchers; designs that reduce the opportunities for satisficing are likely to produce higher quality responses—though may increase non-response if short-cuts are not possible.

1.3.3 Deception and Self-Disclosure Differences

There is evidence that when people lie to one another they do so differently in different media, with different proportions of and different kinds of lies in e-mail, telephone, face-to-face, instant messaging, and so on. For example, participants in a study by Hancock et al. (2004) reported being more likely to lie on the phone than face-to-face, and still less in instant messages and e-mail messages. Lies about people's actions were most common on the phone, and lies about feelings were most common face-to-face; false justifications and reasons were most common in e-mail. A range of differences in what a medium affords (whether there's a reviewable trace that could lead to discovery) and people's purposes in using a medium (small talk versus business) are likely to play a role.

Surveys often ask respondents for sensitive information, and their level of honesty and how much they disclose may well be affected by their disclosure experiences in other communicative situations (see Hancock, Chapter 9 in this volume). Survey modes that create a sense of privacy seem to lead to more honest reporting of information that might embarrass or harm the respondent. Privacy is usually created by removing the interviewer from the data collection process, and so computerized self-administration (e.g., CASI and A-CASI) is now widely used to ask people about their sexual behavior, drug and alcohol use, criminal behavior, and so on. But interviewing systems of the future challenge these traditional distinctions. Consider a text chat interview. This is a mode that clearly involves real-time interaction with another person—an interviewer—but nothing is communicated about the interviewer's appearance or voice. And consider an animated agent interviewer. In this mode, all of the agent's facial and vocal characteristics are evident to the respondent but the agent is clearly not a living human. On intuitive grounds both text chatting and animated agent interviewing seem somewhere between interviewer- and self-administration in terms of the degree to which respondents are likely to truthfully disclose sensitive information, but there isn't empirical evidence yet one way or another.

1.3.4 Social Presence

Face-to-face communication and remote video have high social presence (Short, William, and Christie, 1976) because they allow the exchange of rich interpersonal cues. In comparison, text has low social presence. Thus, we would expect interviewing systems that use video to mediate between respondents and interviewers or that display an animated computer face in a web browser to elicit a greater sense of social presence than a text chat interview. However, people sometimes attribute human-like characteristics to very *inanimate* user interfaces, even line drawings of people, reacting socially and emotionally (Reeves and Nass, 1996). So it could be that interfaces

that communicate any animacy at all, including live text and animated interviewing agents, will create enough social presence that respondents will shade the truth as they sometimes do when human interviewers ask sensitive questions.

How the literature on social presence of different media applies to the survey response task is not yet clear. Although Reeves and Nass (1996) observed social responses were produced by line drawings in a nonsurvey task (rating personality), Tourangeau, Miller, and Steiger (2003) found very little effect of including a still photograph of the researcher in a web survey interface. Perhaps survey respondents really do know the difference between animate and inanimate interface objects and when it is to their advantage to interact with an interviewing system as just a piece of technology—because, for example, they wish to be truthful if the costs are not too great—then they are able to do so. More evidence that survey respondents can ignore or compartmentalize social cues when it is to their advantage to do so is the apparent sense of privacy that is created by listening to a recording of an interviewer’s question via headphones, even though it adds voice—a particularly human attribute—to the text of conventional self-administration. Respondents must understand at a fundamental level that a recorded voice is inanimate.

1.3.5 User Modeling and Adaptive Interfaces

An intense and controversial area of activity in communication technologies has involved “user modeling”—designing interfaces that are tailored to different cultures, groups, or individuals (e.g., see Fink and Kobsa, 2000; Fink et al., 1998; Kay, 1995; Kobsa, 1990, 2001; Rich, 1983, among many others).

In general, there is agreement that making an interface seamless and easy to use is paramount. Where there is disagreement on whether a one-size-fits-all approach can sufficiently satisfy all users, and whether explicitly modeling individual users or groups of users (e.g., experts versus novices) is effective in all domains. The evidence is mixed and varies for different applications and task domains. But there are domains in which user modeling seems to work quite well. One example is intelligent tutoring systems that maintain individual student models, presenting different exercises for different students to work on depending on which skills need practice (e.g., see Corbett et al., 1997). Another is Rich’s (1979) literature recommendation system, which demonstrated that a model of a group of users (or “stereotype model”) can improve the system’s performance and the user’s experience relative to a system that uses no model. The system recommends novels to users that it believes they might like based on descriptions of the novels and the stereotype to which users belong, based on very brief self-descriptions. Users rated the system’s recommendations as far more successful when they were based on the stereotype model than when they were made generically.

For survey technologies, an additional twist is the longstanding and deeply held view that a survey is standardized only if every single respondent experiences precisely the same stimulus (cf. Fowler and Mangione, 1990). So a debate on tailored interfaces must consider what counts as “the same stimulus”—must it look the same and use the same wording? Or, in the interest of standardizing the user/respondent’s interpretation

of survey questions, should interfaces be individualized? We will return to this issue, which we see as at the heart of the conversation between survey researchers and communication technologists, at the end of the chapter.

1.3.6 Exploiting Mode Differences

A final point that survey designers should consider as they approach the scholarship on communication technologies: communication technologists have tended to exploit the information available in different media, and (as a general trend) to create technologies that make use of all available channels of information (e.g., consider measures of vocal stress in automated speech interfaces that can diagnose user frustration as discussed by Person, D’Mello, and Olney in Chapter 10 in this volume). In contrast, the survey world has differing tendencies in thinking about “mode effects” (differences in the quality of data collected, e.g., on the telephone versus face-to-face versus web surveys). One tendency is to see mode effects as undesirable irritants that must be eliminated, as evidenced in the years of concern during the transition from face-to-face to telephone interviewing. Another has been to exploit what a new mode has to offer so as to maximize data quality.

So survey designers thinking about new technologies have a choice. They can either work hard to keep data quality comparable across modes, even if this means that some exploitable features of a mode won’t be used. (For example, assuming respondents’ facial information is useful to diagnosing need for clarification, this would mean *not* using facial information if a survey is also being administered on an audio-only telephone.) The alternative choice is to exploit different modes to maximize data quality, even if this leads to noncomparable data across different modes of administration. Getting used to consciously choosing which strategy to take is one of the challenges that survey designers will face as they consider adopting new communication technologies.

1.4 WHAT COMMUNICATION TECHNOLOGISTS NEED TO KNOW ABOUT SURVEY INTERVIEWS

The survey enterprise is vast and far more complex than most communication technologists are likely to realize. An astonishing amount of thought and care go into question design, sampling, interviewer training and monitoring, so that the resulting data are as unbiased and accurate as possible. This thought and care have resulted in a specialized vocabulary and a set of acronyms² that can be mysterious to the uninitiated. Communication technologists who want to collaborate with survey designers—and the size

²To give a flavor of a few of the acronyms, survey researchers distinguish between CATI (Computer-Assisted Telephone Interviewing—human interviewer on telephone is prompted by computer and enters answers into computer), CAPI (Computer-Assisted Personal Interviewing—human interviewer face-to-face is prompted by computer and enters answers into computer), SAQ (Self-Administered Questionnaire—respondent fills out paper form with interviewer present), A-CASI (Audio-Computer-Assisted Self-Interviewing—respondent hears questions over headphones and enters into laptop, with interviewer

and impact of the survey arena should make this an attractive proposition—will need to become familiar with some basic distinctions.

In thinking through whether to adopt a new communication technology for interviewing, survey designers will want to be confident that if they adopt it:

- Data quality (the quality of the estimates about the population that result from the interviews) is no worse with a new than existing technology. Data quality is measured by a number of indicators. This includes data reliability (the extent to which you get the same estimates when you redo the survey, or compare the answers of similar subgroups in the sample), data validity (the extent to which you are measuring what you claim to be measuring), coverage of the population (the extent to which the sample frame—the list of names or addresses from which the sample is selected—reflects the full population), and nonresponse rates (see below).
- Costs for developing the survey, administering it, training interviewers (if there are any), and data analysis are no greater, or at least remain in the realm of possibility.
- Respondents find the interaction no more irritating or difficult, that is, *respondent burden* of various kinds doesn't increase.
- Respondents are no less likely to agree to participate (*response rates* do not decrease), no more likely to hang up or abort their participation (*completion rates* do not decrease, *break-off rates* do not increase), and no more or less likely to participate or complete the survey if they come from different subpopulations (which would lead to *nonresponse error* if different subpopulations answer differently).

Here is a very rudimentary and oversimplified set of basic facts and distinctions about the survey process and what survey designers worry about along the way. Obviously the list isn't exhaustive, and the distinctions aren't exactly as clear as we're making them sound. A number of reader-friendly introductions to surveys and interviewing present these issues with more precision, including Beimer and Lyberg (2003) and Groves et al. (2004).

1.4.1 Major Distinctions

A primary distinction in survey research concerns whether the respondents report about themselves and the other people with whom they live (*household surveys*) or the organizations to which they belong, usually companies but also schools, farms, hospitals, prisons, and so on (*establishment surveys*). Household surveys are typically

present), and PAPI (Paper and Pencil Interviewing—human interviewer, face-to-face, enters answers on paper). The acronyms result from different historical moments and trajectories rather than a unified labeling scheme, and can thus be less than transparent. Readers who want to familiarize themselves with these terms are recommended to consult the definitions provided in Couper et al. (1998) and to some degree in Chapter 3 in this volume.

quite different from establishment surveys in many ways including the interviewing techniques. For example, the principles of standardized interviewing are far less prevalent in establishment than household surveys, and the responses come from records like the payroll data base in contrast to household survey where they tend to be self-reports (e.g., based on respondent's memory). The particular technologies that might be feasible in the two kinds of surveys are likely to differ not only because of differences in interviewing and the types of data but also because businesses are likely to adopt new technologies sooner than households. For example, web surveys were used regularly in some establishment surveys well before they were used in household surveys primarily because businesses were likely to have high speed Internet access and computers on all desktops prior to households.

Another major distinction is often drawn between surveys that interview respondents just once or on multiple occasions. The former is a *cross-sectional* survey (it takes measures at one slice in time); the latter is a type of *longitudinal* survey (it enables comparisons not only between respondents but within respondents over time). In this kind of longitudinal survey, respondents are considered to be in a *panel*, a group of people brought together to provide data across multiple waves of interviewing. Panel surveys may create unique opportunities to use interviewing technologies that adapt to respondents. Across waves, quite a lot of information becomes known about respondents. For example, if a respondent's attributes, e.g. race or gender or health status, become known in one interview, an interviewing agent might be configured to resemble (or not) the respondent in subsequent interviews.

Yet another major distinction concerns the kinds of questions asked of respondents, specifically whether they are asked for their opinions or about their behaviors. The process of answering questions is similar across these distinctions (see Tourangeau et al., 2000) but there are some differences that could affect the kinds of interviewing technologies one would develop. For example, the answers to behavioral questions can in principle be verified; there is a right or wrong answer to a question about whether one voted in the election. But answers to opinion questions, like how well elected officials are doing at their jobs, really can't be verified. Thus, one can imagine an interviewing technology that is able to confirm certain answers in real time by comparing against online databases (see Marx, Chapter 13 in this volume, for a discussion about the ethics of such technologies). But it's hard to imagine an analogous procedure when the survey concerns opinions.

A final distinction that communication technologists should be aware of is that between closed- and open-ended questions. Closed questions provide a limited number of response options and require respondents to select from the set, as in requiring a respondent to select "yes" or "no"; requiring a number for answers to numerical questions (e.g., number of people living in your house); or requiring answers on continuous or discrete rating scales (e.g., answering on a scale of 1 to 5 where 1 means "strongly agree" and 5 means "strongly disagree"). Open questions give respondents free rein in answering; what they report verbally is usually transcribed by interviewers (in paper self-administered questionnaires, handwritten open text is clerically entered

into the database, and in web-based questionnaires the text is directly entered by the respondent). The text is coded, usually by people with help from computers, so similar responses can be aggregated and tallied. Closed format questions are far more common than open responses primarily because of the extra processing effort associated with open text.

1.4.2 Major Players

A basic issue for communication technologists to be aware of is that the content of a survey, and the politics and funding of its sponsoring agency, can affect how a communication technology gets adopted. A major arena of survey research is government surveys that produce official statistics in domains like employment, crime, health, business, energy, and demographics. Governments publish official statistics on a cyclical basis (e.g., monthly unemployment statistics) and so rely on well established administrative and operational machinery to collect and analyze survey data. Official statistics tend to be based on answers to behavioral questions, but occasionally opinion questions are asked in government surveys. Some official statistics are produced on an international scale, such as the Eurobarometer, which involve collaboration between government agencies, designed to allow cross-national comparisons.

Another set of scientific surveys, measuring both opinions and behaviors, often over time, are carried out by academic survey centers affiliated with universities. Large-scale academic surveys share much with government surveys but differ in important ways. One key difference is that academic surveys are typically designed to produce data that will allow social scientists to test theories with multivariate methods. As a result, such surveys often present batteries of related questions to respondents so that the researchers can construct scales by combining answers. It is also the case that academic surveys and their sponsors do not have the deep pockets of many national government survey organizations and so may be less able to bear the financial costs of adopting new technologies than would a government agency.

Another important sector in the survey research industry—in some ways the most visible face of the industry—is the corporate entities that specialize in measuring public opinion. These surveys obviously ask questions that are primarily about opinions, but they also collect information about respondents' backgrounds, for example, about their education and income. Finally, market research involves vast amounts of survey research. This work is often done on very tight schedules and one-time surveys are the norm. The lion's share of market research is now done online, in contrast to how it is conducted in the other sectors we have mentioned.

The costs of adopting a new technology are greater for long-running, large-scale surveys than for smaller scale and less entrenched surveys. Researchers are reluctant to change the administrative machinery that supports ongoing surveys and are loath to compromise time series by changing the mode of data collection. In favor of adopting

new technology is that the most high-impact surveys have the most to lose by failing to keep up with new communication technologies. Changes in how the population communicates will require increased use of mobile phones, the web, and chat methods in order to reach respondents.

In any case, understanding the administrative and financial obstacles to expanding how a survey does its business will be essential for communication technologists interested in collaborating with survey researchers. The incentives and obstacles to implementing new communication technologies for survey interviewing are likely to vary in these different settings.

1.5 WHAT HAPPENS IN SURVEY INTERVIEWS

1.5.1 Sequence of Events

The sequence of events in an interview actually includes a few steps before the questions are asked. First, there is a contact initiation phase—a telephone call to a household, an e-mail solicitation, a postal mail request. In a telephone call to a household, a respondent must be selected from among possible household members; not all respondents are interchangeable or even appropriate for particular kinds of questions. An important task for interviewers is the process of what is called *refusal conversion*: convincing reluctant respondents to participate. This isn't done to annoy respondents, but rather because the sampling procedures that have led to the selection of the household and respondent require that a reasonable percentage of the selected respondents participate to increase the chances that the resulting population estimates are accurate.³ After obtaining agreement to participate, interviewers work to present standardized interviews and maintain rapport such that the interview is appropriately completed.

Note that each of these steps—contact initiation, respondent selection, refusal conversion, and interview completion—can look different with different modes of interviewing. For example, all of the activities until the first question is asked are largely unscripted, but once the interview begins, interviewers typically are far more limited in the words they are empowered to use. Thus, technologies used in the early stages of interaction will need to be capable of supporting exchanges in which the particular words and conversational turns have not been planned, whereas the technology that supports the actual interview may not need to be so flexible, depending on the interviewing approach. Any new communication technology that is considered for survey interviews will have to be able to allow each of these steps to be carried out adequately, and this may require multiple technologies for the different stages.

³ Survey organizations expend the effort to persuade initial non-respondents to participate because higher response rates are often assumed to be associated with lower non-response bias, i.e. the difference between the answers of respondents and non-respondents when the answers of the latter group are somehow knowable. In fact response rates seem not to be a good predictor of non-response bias (e.g. Keeter, et al., 2000). Yet refusal conversion continues to be a key activity of survey interviewers.

1.5.2 Features of Interviews

Interviews themselves are quite an unusual form of interaction when compared with other kinds of conversations, and self-administered surveys are different from other kinds of interactions with computers. Their distinctive features will need to be considered by communication technologists working on survey applications. First, in interviewer-administered surveys, the two parties in the interaction are in unusual roles (e.g., see, Clark and Schober, 1991; Schaeffer, 2002; Schober, 1999; Schober and Conrad, 2002; Suchman and Jordan, 1990) compared with roles in question-asking in spontaneous conversations. Standardized interviewers follow a script of questions and probes written by others, and so they are really acting as intermediaries for the survey designers who wrote the questions. When an interviewer reads a question, she isn't formulating it with the same freedom and responsivity to prior talk that spontaneous conversationalists engage in; when she refuses to explain what a question means (as she should if she is carrying out a strictly standardized interview), she is behaving in ways that would be at least impolite, if not bizarre, in everyday interactions (and which potentially harm data quality: see Conrad and Schober, 2000; Suchman and Jordan, 1990; Schober and Conrad, 1997). Whether or not respondents know it, there is an invisible third agent—the survey designer—involved. This can affect respondents' cognitive processes and emotional reactions to the interaction in odd ways: if respondents fail to recognize that the interviewer's behavior is constrained, they can interpret the interviewer's robotic behavior as uncooperative or even hostile, and they can interpret the interviewer's scripted "okay" as denoting real approval when it may not (Schober and Conrad, 2002). Savvy respondents who understand the game are operating by somewhat different conversational rules than they would be under other circumstances.

Computer-assisted interviews add an extra element to the situation (see Couper, Chapter 3 in this volume). Unlike a piece of paper with only static instructions to the interviewer to read the next question, a computer program is active and potentially leads the interviewer down different paths depending on respondents' prior answers (e.g., skipping all the questions about job satisfaction if the respondent reported being unemployed). There is thus an additional attention-demanding agent in the situation, or, to put it another way, the survey designer structures (intrudes into?) the interview in a new way through computer-assisted interviews.

1.5.3 Features of Self-administered Surveys

As interviews differ from other kinds of human conversation, interactions with self-administered survey systems are different from other kinds of human-computer interaction (Schober, Conrad, Ehlen, and Fricker, 2003). For one, the system initiates the interaction rather than the user; more often computer users go online to obtain information, but in online surveys they provide it. The location of the information being transferred is different: it comes from the user's memory rather than being something that can be found in existing documents. And, as we have demonstrated (Schober, Conrad, Ehlen, and Fricker, 2003), respondents to web surveys are more

likely to (mistakenly) assume that they understand what words in surveys mean than to assume that their search queries are interpretable by a search engine.

This is not to suggest that online textual surveys are entirely different from other kinds of human–computer interaction, but simply that they differ in important and systematic ways that may be useful to consider when thinking through new communication technologies for surveys.

We should note that new technologies for surveys that blur the distinction between interacting with an interviewer and interacting with a computer are blurring the distinction between CMC (computer-mediated communication between two people) and HCI (human–computer interaction). Agent-like interviewing systems, for example, already have some features of human interviewers (spoken voice, dialogue responsiveness, facial movement), and it is not yet clear whether principles and experiences from HCI or CMC are the more appropriate to bring to bear.

1.6 COMPARABILITY OF QUESTIONS ACROSS MODES

Ideally, survey questions are designed to produce comparable answers no matter what the medium of communication. Years of testing compared the answers produced via landline telephone with those produced face to face, with the general consensus that the answers were sufficiently comparable to adopt the new medium (see de Leeuw and van der Zouwen, 1988). On the other hand, we now know that self-administered interviews can lead to greater (presumably more honest) reporting of sensitive behaviors than face-to-face interviews (Tourangeau & Smith, 1996). So comparability of questions across modes—where comparability involves wording as well as technological and social context—is important for communication technologists to think about.

Considering that the mode of administration can affect the comparability of questions, there are some dangers in adopting a new technology for administering an existing questionnaire. For example, survey researchers often pretest questions to identify problems in interviewer delivery and respondent answering, but if a questionnaire is inherited from an earlier survey in a different mode, it is not clear that problems identified earlier necessarily recur with the new technology and it is also possible that new problems will occur when the new technology is introduced. This will depend on how similar the old and new modes are in these *affordances* (see page 8.) For example, moving from telephone to IM means that there is a reviewable record of the earlier exchanges between interviewer and respondent and an opportunity to reread questions that might have been too long to grasp in a single interviewer reading over the phone; but there is still the temporal pressure of a real-time interview, which could limit “the respondent’s likelihood of” reviewing text in contrast to a self-administered questionnaire on the web.

It remains to be seen whether new media for survey interviewing will require new kinds of pretesting, or whether the kinds of problems uncovered via today’s means will be observed with the some pretesting methods.

1.7 CONTENT AND FORMAT OF QUESTIONS

The impact of a new communication technology may depend on both the content and format of questions.

1.7.1 Content

The opinions that respondents report can be affected by interviewers' observable (or inferable) attributes (e.g., race, gender, age) if these are relevant to the question topic. This general phenomenon is well illustrated by a survey of black households (Schuman and Converse, 1971) in which either white or black interviewers asked respondents a variety of questions. On many questions, there was virtually no difference due to the interviewers' race. When the questions concerned race, however, respondents expressed different opinions to black than white interviewers. For example when a black interviewer asked these black respondents if they could trust white people, 7% said "yes." When white interviewers asked the same question, 35% said "yes." These sorts of findings suggest that new communication technologies that highlight an interviewer's (or interviewing agent's) race, gender, or age group could affect answers to questions about these topics differently from technologies that do not (although it is also not clear whether one can devise truly neutral interviewing agents!). Cassell and Miller (Chapter 8 in this volume) discuss this issue in more detail.

As we have noted earlier, another way that impact of question content may differ by interviewing technology concerns sensitive questions. If a new technology reduces people's comfort enough that they stop reporting embarrassing behaviors, survey researchers will be wary of adopting it. Thus, the social presence afforded by different media will be a key consideration in the use of new media: streaming video seems likely to produce effects similar to what is observed in face-to-face interviews, but it is unclear to what degree virtual animated interviewers, IM, or speech dialogue systems create a sense of social presence for respondents.

1.7.2 Format

The format of response options to questions can affect the kinds of answers that respondents give, and these different formats will be differently possible to implement in different media. Closed options are already implemented quite differently in today's technologies; when an interviewer or a recorded voice speaks the response options, depending on the implementation the respondent must either listen to the entire list or can interrupt before the full set of options is listed. In a text-based self-administered interview, the respondent can see all the options at once and may even consider them before reading the question (or may ignore them—see Graesser et al., 2006).

Closed responses have been the norm in today's surveys, as they are far easier to capture and analyze than open-ended answers. But new communication technologies are opening up possibilities for systematic analysis of open-ended answers that go far beyond what is currently possible. The AutoTutor program, for example, uses Latent Semantic Analysis to allow students in tutoring situations to type natural-language

answers, and it then classifies responses, with notable accuracy, so as to engage in fairly realistic dialogue. As such technologies improve, there is no reason to imagine that they won't be implementable in survey situations for dealing with written and spoken answers in far more naturalistic ways.

1.8 ERROR IN SURVEYS

As noted earlier, survey researchers will be unwilling to adopt a new communication technology if it is shown to increase survey error. There are a number of extensive treatments of different kinds of survey error (Groves, 1989), but the main areas that need to be considered include the following:

Coverage error. Coverage error derives from a mismatch between those people or organizations in the sampling frame—the list of addresses, phone numbers, email addresses that represents the population to be surveyed—and those people or organizations who are actually in the population. This mismatch produces coverage error when those not in the frame differ on the attributes of interest from those in the frame. For example, if the sampling frame does not include people whose only telephone is a mobile device and these mobile-only users differ on the attributes being measured (e.g. political preferences) this would produce coverage error. The larger the undercoverage (the number of cell-only respondents in the population), the greater the total impact of coverage error on the survey results.

Sampling Error. Sampling error occurs when the people one selects from a sampling frame don't accurately represent the population from which they are selected (e.g., report different opinions and behaviors than would the entire population if everyone were interviewed) and to which one wants to generalize. No sample will ever be a perfect microcosm of the population and different samples will almost surely lead to slightly different results; a certain amount of error is just inherent in the sampling process. In general, the larger the sample size the smaller the sampling error.

There is a vast amount of theoretical and practical work aimed at understanding and reducing sampling error. Because the current volume is concerned with collecting data from respondents after they have been selected in a sample, this source of error is unlikely to interact with the choice of interviewing technology. However, a communication researcher working in the survey domain needs to recognize that survey designers will be concerned about how adopting the technology affects the sampling.

Nonresponse error. Nonresponse error occurs when the people in the sample who didn't participate in the survey differ in the opinions or behaviors being measured than those who did participate. If a new communication technology selectively irritates or scares or confuses certain subgroups in the sample enough that they don't participate in or finish the survey, and they would have answered the questions differently than the groups with higher participation and

completion rates, this could lead to nonresponse error. As with coverage error, the overall impact of this discrepancy depends on the nonresponse rate. Because interviewers are involved in recruiting respondents and keeping them on task, this type of error is quite relevant to the choice of interviewing technology.

Measurement Error. Measurement error is how far off from true values respondents' answer are. Measurement error can result from *interviewer-related* error—e.g. bias due to particular interviewers' behavior or observable attributes—or *respondent-related* error—e.g. problems in respondents' comprehension, memory, or judgment that affect answers. There is some evidence that variation in interviewer “probing” behavior (e.g., not always asking the respondent to select from the response categories) increases the impact of interviewers on answers (Mangione, Fowler, and Louis, 1992). Such variation might be substantially reduced by a speech dialogue system that is clever enough to recognize when a probe is required and administers the probes systematically. Of course, the particular voice and dialect chosen by the designers can introduce error of its own. Evidence of respondent-related measurement error would be lower reports of embarrassing behaviors with a new technology, presumably because the technology introduces social presence that was not present in earlier technology.

It is worth elaborating on respondent-related error, given the rather large survey literature giving examples (often amusing, but with disturbing consequences) of what can influence answers (e.g., see Clark and Schober, 1991; Schuman and Presser, 1981; Sudman, Bradburn, and Schwarz, 1996; Tourangeau et al., 2000, among others). Seemingly minor changes in wording, in question order, and in response alternatives can (though they don't always) dramatically affect the percentages of respondents providing particular answers. These phenomena are known as *response effects*, and they can occur at a number of different levels.

To give a flavor, consider Loftus's (1975) comparison of answers to these two questions:

1. Do you get headaches frequently, and if so, how often?
2. Do you get headaches occasionally, and if so, how often?

Respondents to the first question reported an average of 2.2 headaches per week, while those for the second reported an average of 0.7 headaches per week. This presumably results from what kinds of headaches respondents consider they are being asked about (see Clark and Schober, 1991).

Or consider the study by Schwarz, Hippler, Deutsch, and Strack (1986) which asked Germans about their television-watching habits. When asked how many hours of television they watched each week, respondents given a set of response alternatives ranging from “less than 1/2 hour” to “more than 2 1/2 hours” provided notably lower estimates than respondents given a set of alternatives ranging from “less than 2 1/2 hours” to “more than 4 1/2 hours.” Extrapolating from these data to the population would have provided an estimate of 16% of the population watching more than 2 1/2

hours of television per week with the first response scale, but 38% of the population with the second response scale. In general, the nature of response alternatives (Is there a “don’t know” option? Must respondents choose from existing alternatives, or should they provide open-ended responses?) dramatically affect in what answers respondents give.

So far, the story on whether these response effects work the same way with new interviewing technologies isn’t entirely clear. Comparisons of current modes reveal some response effects that occur both in interviews and self-administered questionnaires as well as self-administered questionnaires. So, for example, just as in telephone interviews, respondents to self-administered paper and pencil questionnaires are more likely to answer that they “don’t know” when a question’s response options explicitly include “don’t know” (Schuman and Presser, 1981).

On the other hand, some response effects definitely work differently in different interviewing media. For example, recall that when an interviewer presents response options, respondents are more likely to go with the most recent one. But when they answer the same question on paper, they are more likely to go with the first response option (Krosnick and Alwin, 1987). As another example (Schwarz, Strack and Mai 1991; Strack, Martin and Schwarz, 1987), question order in interviews can make a big difference: people who first report how satisfied they are with their romantic life and then answer about how happy they are in general tend to answer in the same direction, i.e. their answers are correlated; their answers are uncorrelated when the questions are asked in the reverse order (general and then specific). But on paper (self-administered), this order effect is attenuated, presumably because people can look back and forth and change their answers.

As we see it, the moral is that new communication technologies for surveys will need to make sure they don’t create new kinds of response effects.

1.9 ADOPTING NEW TECHNOLOGIES

Our assumption is that the communication technologies emerging today are the tip of the iceberg, and that the case studies in this volume may not be timely in future. Nonetheless, we believe that considering these examples helps raise sets of questions that will continue to be important as new technologies are developed and become options for use in survey interviewing.

Adopting current and upcoming communication technologies for survey interviewing could lead to a number of benefits, assuming the technologies have the right properties. One could imagine reducing interviewer-related error by making entirely standardized interviewing agents (although this isn’t quite as straightforward as it seems). There are certainly potential cost savings from automation, e.g., in not hiring as many human interviewers, or increasing the interviewing pool by making interviewers less location-bound. Multimodal technologies might increase accessibility for respondents with sensory and motor deficits, or for people who are immobile or unreachable. Collecting paradata from respondents might allow better interpretation of responses and better diagnosis of respondents’ needs.

But the potential benefits for any particular technology may just as easily be outweighed by the downsides, and the ones that come to mind are not trivial. As we have collected the work presented in this volume, a number of practical and ethical questions, to which we do not have answers, have come to the fore. As we see it, survey researchers need principles to guide decision-making about adopting new communication technologies in surveys in at least the following areas:

1. Should a new technology only be adopted for a survey once enough people in the population being measured have access to it? Once enough people use it frequently in their own communication, and so are used to using it? What counts as enough?
2. What are the costs of *not* adopting new technologies? For example, will a survey using an old technology (e.g., paper, or face-to-face interviews), or not allowing alternative modes for answering, seem antiquated or dull or nonscientific?
3. How connected does the new technology have to be to what people already know? WebTV assumes people know how to use a TV; agent technology assumes people already know how to interact face to face; IM assumes respondents can key in text. Just how different can a technology be from what a respondent already knows to allow accurate survey measurement (with all this entails—appropriate coverage, response rates, completion, satisfaction, validity)?
4. Is there a gold standard to which all new technologies should be compared? Is a face-to-face interview administered by an experienced personable interviewer the best, and a new technology should only be adopted to the extent that the interaction—or, to complicate matters, the data quality—mimics what happens in that sort of interview? Or might mediated communication potentially lead to improved data quality in some domains by reducing the social presence produced by a live interviewer as with A-CASI?
5. Survey designers have tried to increase rapport between interviewers and respondents. Communication technologists have tried to increase usability with computer interfaces. As survey interfaces have the potential to become more anthropomorphic, does rapport become part of usability? Should it?
6. How do we conduct interviews when more and more members of a population don't share the same first language, or the same set of cultural beliefs? How are interviews different when interviewers in Bangalore, rather than Michigan, are calling U.S. households? Should a remote video-mediated or animated survey interviewer look and sound and dress like the respondent? Should a web interface tailor interviews to local background cultural beliefs (e.g., about whether the number of children you report having includes ones who didn't survive)?
7. If new technologies allow interviews to be tailored to characteristics of individual respondents, will this infringe upon privacy, or is such infringement a necessary feature of a society that collects data about its citizens? Will respondents' privacy have been violated if survey questions are tailored to what is known about them in databases from prior survey responses, monitored web use, or linked datasets? The flip side: If you know something about a respondent

and don't divulge that, or don't rely on it as you ask questions, is it ethical to waste their time asking again? Is it ethical not to confirm and to just assume from what you collected via surveillance?

8. New communication technologies increase the number of ways that respondents can be contacted and surveilled and surveyed; along with this the opportunities to hassle and burden people increase. What level of intrusion is acceptable?
9. Informed consent: Is it ethical to capture paradata (e.g., typing errors, changed answers, response times, speech disfluencies, facial expressions) without explicit consent? When people agree to participate are they aware that this is being captured? Should they be?
10. There is an obvious technical survey coverage issue when not everyone has access to the same technologies. Is there an ethical issue, above and beyond this, about haves and have-nots? Or is it fine for some respondents to have access to technologies that others don't as long as everyone is able to participate somehow?

1.10 THE THEORETICAL HORIZON

The issues raised by contemplating new communication technologies for survey interviewing are serious in both practical and political terms, as sorting out what comes next will affect how we think of the role of consent in data-gathering about citizens, and ultimately what kinds of societies we become.

The issues also raise some basic theoretical questions about the nature of interaction and about standardization, and the answers are not straightforward.

1.10.1 Interaction

As we see it, the new communication technologies currently on the horizon all fall somewhere on the continuum between face-to-face interaction and written asynchronous communication. Videomediation keeps most of the features of face-to-face interaction while removing physical copresence; textual web interfaces like IM add a back-and-forth turn-taking structure into the "dialogue" that could allow clarification dialogue (Conrad, Schober, and Coiner, 2007); survey systems that diagnose facial paradata from respondents are emulating parts of what human interviewers can do. The same is also true for features of interviewers/interviewing agents: interviewers have a number of social identities, some of which are differently evident or inferable depending on the medium of communications. For example, accent cues are unavailable under textual mediation but can lead to all sorts of attributions about interviewers on the phone, and appearance, dress, and facial cues are unavailable on the phone but can lead to all sorts of attributions face to face. Thus, the range of behavioral and social cues from interviewers (and, in the other direction, respondents) give degrees of anthropomorphism to the interfaces (Schober and Conrad, 2007), with variability on the dimensions of dialogue capability, perceptual realism, and intentionality.

What is new, both practically and theoretically, is that interviewing systems will not only allow but require *conscious choice* about which elements of anthropomorphism are implemented. If one considers videomediated interviewing, then visible characteristics of the interviewer are of necessity part of the equation, with all the attendant benefits and baggage. If one considers building an automated animated agent, that agent must have an appearance and resulting inferable social identity; the survey researcher will now have to *choose* whether to create only one interviewer with one appearance for all respondents, or to create alternate interviewers for different respondents, or to allow respondents to choose their interviewer's appearance. One can imagine changing the appearance and voice of an interviewer mid-interview once the topic of questioning changes; one can imagine changing the interviewing agent's responsivity for different questions, with friendly "okays" for nonsensitive questions, but much more measured responses that don't suggest approval or disapproval for sensitive questions. One can imagine changing the interviewer's responsivity depending on the system's measurement of the respondent's reactions as the interview goes along; for example, if the respondent becomes irritated or agitated the system might loosen some of the constraints of standardization.

Having this kind of choice at our disposal is something brand new. It means that aspects of interviewer-related influence that have always been assumed to be error in the data may not only be controllable, but even exploited, if it turns out that this is helpful. It also means that survey designers will need to have far greater knowledge about interactive processes from basic social sciences, perhaps even more than the social scientists currently know. Questions that survey researchers need to know the answers to could thus drive new basic research in the social sciences, with the outcome that we will end up understanding where survey interviews of different kinds fit into a grander taxonomy of tasks.

To complicate things even further, new technologies may allow for access to information about respondents that goes beyond what face-to-face interviewers have available to them. Automated analysis of respondents' language, for example, could potentially let an interviewer, or an interviewing system, know when a respondent is likely to be under- or overreporting a behavior, and the system could intervene to probe further. We fully recognize the disturbing implications of following this line of thinking, and do not mean to suggest that we advocate building Brave New World systems of this sort. But given that technologies for detecting and augmenting evidence of one's partner's mental and physical processes are in development for use in other arenas [we are thinking of "augmented reality" systems for benign purposes like having beyond-visibility access to, say, your chamber music collaborator's breathing or bowing (Schober, 2006)], the space of technological applications to surveys may range beyond the continuum from face-to-face to text.

1.10.2 Standardization

The facts of new interviewing technologies complicate survey researchers' notions of standardization to a new degree. Current standardized methods arose in reaction to a

long history of missteps and errors, in response to evidence that different interviewers were affecting the quality of answers differentially (see Beatty, 1995). The solution was to standardize question wording and interviewer delivery of those questions, such that all respondents receive precisely the same wording, and interpretation of questions is left entirely to respondents.

We have argued elsewhere, along with others (e.g., Conrad and Schober, 2000; Schober and Conrad, 1997, 2002; Suchman and Jordan, 1990), that although the goals of standardization are laudable, the logic of this solution is problematic. As Suchman and Jordan put it, it is perhaps more consistent with the goals of standardization to standardize the *meaning* of each question for respondents, rather than the wording, even if this leads to different respondents being presented with different words. So, for example, if a respondent has a different interpretation of what it means to live in a household than the government agency asking the question, it might be useful for the interview to allow the kind of idiosyncratic interaction that would allow this discrepancy to be uncovered and for the respondent's report of how many people live in his house to conform with other respondents' interpretation. We have shown that data quality can be improved, although interviews take longer, on the telephone (Conrad and Schober, 2000; Schober and Conrad, 1997; Schober, Conrad, and Fricker, 2004), face to face (Schober et al., 2006), and in automated textual (Conrad, Schober, and Coiner, 2007) and speech (Ehlen, Schober, and Conrad, 2007; Schober et al., 2000) interviewing systems that allow nonstandardized clarification dialogue to take place.

New interviewing technologies that allow survey researchers to choose features of anthropomorphism complicate the logic even more. Following the logic of strict standardization, a survey interview of the future will be most standardized if all the available features of the interviewer or interviewing system are exactly the same: if it looks the same, sounds the same, and responds the same. Thus, any variation in answers will not be attributable to variation in the interviewer. (This makes such an interviewing system a better approximation of strict standardization than is possible in any current call center.) But extrapolating from the findings on standardizing wording versus meaning, perhaps what is more desirable is that *data quality* be comparable across all respondents. By this logic, any aspect of an interviewer or interviewing agent that potentially gets in the way of data quality—say, race of interviewer for certain subpopulations for certain kinds of questions, responsiveness of the interviewer on sensitive questions—should be tailored for the particular respondent population, to meet the goal of standardizing the respondent's *experience* of the interview. So whatever it takes for the interview to feel nonthreatening, while asking embarrassing questions, should be implemented, and if that means making the interviewer look and sound motherly for one sort of respondent but very distant and computer-like for another sort, then that will satisfy the goals of standardization.

Current data do not allow advocacy of one position over the other, or of another position entirely. But this is the sort of very basic question that thinking about new communication technologies for surveys will require us to deal with.

REFERENCES

- Anderson, A. H., et al. (1997). The impact of VMC on collaborative problem-solving: an analysis of task performance, communication process and user satisfaction. In K. Finn, A. Sellen, and S. Wilbur (Eds.), *Video-Mediated Communication* (pp. 133–155). Mahwah, NJ: Lawrence Erlbaum, Inc.
- Beatty, P. (1995). Understanding the standardized/non-standardized interviewing controversy. *Journal of Official Statistics*, 11, 147–160.
- Biemer, P. P., and Lyberg, L. E. (2003), *Introduction to Survey Quality*, New York: Wiley.
- Clark, H. H. (1992). *Arenas of Language Use* (pp. 176–197). Chicago: University of Chicago Press.
- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Clark, H. H., and Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, and S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 127–149). Washington, DC: APA.
- Clark, H. H. and Schober, M. F. (1991). Asking questions and influencing answers. In J. M. Tanur (Ed.), *Questions About Questions: Inquiries into the Cognitive Bases of Surveys* (pp. 15–48). New York: Russell Sage Foundation.
- Conrad, F. G. and Schober, M. F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly*, 64, 1–28.
- Conrad, F. G., Couper, M. P., Tourangeau, R., and Peytchev, A. (2006). Use and non-use of clarification features in web surveys. *Journal of Official Statistics*, 22, 245–269.
- Conrad, F. G., Schober, M. F., and Coiner, T. (2007). Bringing features of human dialogue to web surveys. *Applied Cognitive Psychology*, 21, 165–188.
- Conrad, F. G., Schober, M. F., and Dijkstra, W. (2008). Cues of communication difficulty in telephone interviews. In J. M. Lepkowski et al. (Eds.), *Advances in Telephone Survey Methodology* (pp. 212–230). Hoboken, NJ: John Wiley & Sons.
- Corbett, A. T., Koedinger, K. R., and Anderson, J. R. (1997). Intelligent tutoring systems (Chapter 37). In M. G. Helander, T. K. Landauer, and P. Prabhu (Eds.), *Handbook of Human-Computer Interaction*, 2nd ed. Amsterdam, The Netherlands: Elsevier Science.
- Couper, M. P., et al, (Eds.). (1998). *Computer Assisted Survey Information Collection*. Hoboken, NJ: John Wiley & Sons.
- de Leeuw, E. D., and van der Zouwen, J. (1988). Data quality in telephone and face to face surveys: a comparative meta-analysis. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls, and J. Waksberg (Eds.), *Telephone Survey Methodology* (pp. 283–300). Hoboken, NJ: John Wiley & Sons.
- Dillman, D. A. (2000). *Mail and Internet Surveys: The Tailored Design Method*. Hoboken, NJ: John Wiley & Sons.
- Ehlen, P., Schober, M. F., and Conrad, F. G. (2007). Modeling speech disfluency to predict conceptual misalignment in speech survey systems. *Discourse Processes*, 44.
- Ehrlich, J., and Riesman, D. (1961). Age and authority in the interview. *Public Opinion Quarterly*, 24, 99–114.
- Fink, J. and Kobsa, A. (2000). A review and analysis of commercial user modeling servers for personalization on the World Wide Web. *User Modeling and User-Adapted Interaction*, 10(3–4), *Special Issue on Deployed User Modeling*, 209–249.

- Fink, J., Kobsa A., and Nill, A. (1998). Adaptable and adaptive information provision for all users, including disabled and elderly people. *New Review of Hypermedia and Multimedia*, 4, 163–188.
- Fowler, F. J., and Mangione, T. W. (1990). *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park, CA: SAGE Publications.
- Goodwin, C. (1991). *Conversational Organization: Interaction Between Speakers and Hearers*. New York: Academic Press.
- Graesser, A. C., Cai, Z., Louwrese, M., and Daniel, F. (2006). Question understanding aid (QUAID): a web facility that helps survey methodologists improve the comprehensibility of questions. *Public Opinion Quarterly*, 70, 3–22.
- Gray, W. D., and Fu, W. T. (2005). Self-constraints in interactive behavior: the case of ignoring perfect knowledge-in-the-world for imperfect knowledge-in-the-head. *Cognitive Science*, 28, 359–382.
- Groves, R. M. (1989). *Survey Errors and Survey Costs*. Hoboken, NJ: John Wiley & Sons.
- Groves, R. M., and Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J. A. (Eds.). (2002). *Survey Nonresponse*. Hoboken, NJ: John Wiley & Sons.
- Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2004). *Survey Methodology*. New York: Wiley.
- Hancock, J. T., and Dunham, P. J. (2001). Language use in computer-mediated communication: The role of coordination devices. *Discourse Processes*, 31, 91–110.
- Hancock, J. T., Thom-Santelli, J., and Ritchie, T. (2004). Deception and design: the impact of communication technologies on lying behavior. *Proceedings of the Conference on Computer Human Interaction*, Vol. 6, pp. 130–136. New York: ACM.
- Hatchett, S., and Schuman, H. (1975). White respondents and race of interviewer effects. *Public Opinion Quarterly*, 39(4), 523–528.
- Kane, E., and Macaulay, L. (1993). Interviewer gender and gender attitudes. *Public Opinion Quarterly*, 57, 1–28.
- Kay, J. (1995). Vive la difference! Individualized interaction with users. In C. S. Mellish (Ed.), *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (pp. 978–984). San Mateo, CA: Morgan Kaufmann.
- Keeter, S., Miller, C., Kohut, A., Groves, R., and Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly*, 64, 125–148.
- Kobsa, A. (1990). User modeling in dialog systems: potentials and hazards. *AI and Society* 4(3), 214–240.
- Kobsa, A. (2001). Generic user modeling systems. *User Modeling and User-Adapted Interaction*, 11(1–2), 49–63.
- Krauss, R. M., Chen, Y., and Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In M. Zanna (Ed.), *Advances in Experimental Social Psychology* (pp. 389–450). San Diego, CA: Academic Press.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213–236.
- Krosnick, J., and Alwin, D. (1987). An evaluation of a cognitive theory of response order effects in survey measurement. *Public Opinion Quarterly*, 51, 201–219.

REFERENCES

- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, 7, 560–572.
- Mangione, T. W., Fowler, F. J., and Louis, T. A. (1992). Question characteristics and interviewer effects. *Journal of Official Statistics*, 8, 293–307.
- Maynard, D., Houtkoop-Steenstra, H., Schaeffer, N. C., and van der Zouwen. (Eds.). *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*. New York: John Wiley & Sons.
- McNeil, D. (1992). *Hand and Mind. What Gestures Reveal About Thought*. Chicago: University of Chicago Press.
- Prewitt, K. (2006). Keynote address at Second International Conference on Telephone Survey Methodology, Miami, FL.
- Reeves, B., and Nass, C. (1996). *The Media Equation*. Stanford, CA: Center for the Study of Language and Information.
- Rich, E. (1979). User modeling via stereotypes. *Cognitive Science*, 3, 329–354.
- Rich, E. (1983). Users are individuals: individualizing user models. *International Journal of Man-Machine Studies*, 18, 199–214.
- Schaeffer, N. C. (2000). Conversation with a purpose – or conversation? Interaction in the standardized survey interview. In D. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, and van der Zouwen, J. (Eds.), *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*. (pp. 95–124). New York: John Wiley & Sons.
- Schober, M. F. (1999). Making sense of questions: an interactional approach. In M. G. Sirken, D. J. Hermann, S. Schechter, N. Schwarz, J. M. Tanur, and R. Tourangeau (Eds.), *Cognition and Survey Research* (pp. 77–93). Hoboken, NJ: John Wiley & Sons.
- Schober, M. F. (2006). Virtual environments for creative work in collaborative music-making. *Virtual Reality*, 10(2), 85–94.
- Schober, M. F., and Bloom, J. E. (2004). Discourse cues that respondents have misunderstood survey questions. *Discourse Processes*, 38, 287–308.
- Schober, M. F., and Brennan, S. E. (2003). Processes of interactive spoken discourse: the role of the partner. In A. C. Graesser, M. A. Gernsbacher, and S. R. Goldman (Eds.), *Handbook of Discourse Processes* (pp. 123–164). Mahwah, NJ: Lawrence Erlbaum Associates.
- Schober, M. F., and Conrad, F. G. (2002). A collaborative view of standardized survey interviews. In D. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, and J. van der Zouwen. (Eds.), *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview* (pp. 67–94). Hoboken, NJ: John Wiley & Sons.
- Schober, M. F., and Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly*, 61, 576–602.
- Schober, M. F., and Conrad, F. G. (2007). Dialogue capability and perceptual realism in survey interviewing agents. Paper presented at the 62nd Annual Conference of the American Association for Public Opinion Research, Anaheim, CA.
- Schober, M., Conrad, F., Ehlen, P., and Fricker, S. (2003). How web surveys differ from other kinds of user interfaces. *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

- Schober, M. F., Conrad, F. G., Ehlen, P., Lind, L. H., and Coiner, T. (2003). Initiative and clarification in web-based surveys. *Proceedings of American Association for Artificial Intelligence Spring Symposium: Natural Language Generation in Spoken and Written Dialogue* (pp. 125–132). Menlo Park, CA: American Association for Artificial Intelligence.
- Schober, M. F., Conrad, F. G., and Fricker, S. S. (2004). Misunderstanding standardized language in research interviews. *Applied Cognitive Psychology*, 18, 169–188.
- Schuman, H., and Converse, J. M. (1971). The effects of black and white interviewers on white respondents in 1968. *Public Opinion Quarterly*, 35(1), 44–68.
- Schuman, H., and Presser, S. (1981). *Questions and Answers in Attitude Surveys: Experiments in Question Form, Wording, and Context*. New York: Academic Press.
- Schwarz, N., Hippler, J., Deutsch, B., and Strack, F. (1985). Response scales: effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly*, 49, 388–385.
- Schwarz, N., Strack, F., and Mai, H.-P. (1991). Assimilation and contrast effects in part-whole question sequences: a conversational logic analysis. *Public Opinion Quarterly*, 55, 1–25.
- Short, J. A., Williams, E., and Christie, B. (1976). *The social psychology of telecommunications*. New York: John Wiley & Sons.
- Strack, F., Martin, L. L., and Schwarz, N. (1988). Priming and communication: The social determinants of information use in judgments of life-satisfaction. *European Journal of Social Psychology*, 18, 429–42.
- Simon, H. A. (1957). *Models of Man*. Hoboken, NJ: John Wiley & Sons.
- Suchman, L., and Jordan, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association*, 85, 232–241.
- Sudman, S., Bradburn, N., and Schwarz, N. (1996). *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco, CA: Jossey-Bass.
- Tourangeau, R., Couper, M. P., and Steiger, D. M. (2003). Humanizing self-administered surveys: experiments on social presence in web and IVR surveys. *Computers in Human Behavior*, 19, 1–24.
- Tourangeau, R., and Smith, T. W. (1996). Asking sensitive questions: the impact of data collection mode, question format and question content. *Public Opinion Quarterly*, 60, 275–304.
- Tourangeau, R., Rips, L., and Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Whittaker, S. (2003). Theories and methods in mediated communication. In A. C. Graesser, M. A. Gernsbacher, and S. R. Goldman (Eds.), *Handbook of Discourse Processes* (pp. 243–286). Mahwah, NJ: Lawrence Erlbaum, Inc.