Survey Interviews with New Communication Technologies:

Synthesis and Future Opportunities

Arthur C. Graesser, Moongee Jeon, and Bethany McDaniel

The University of Memphis

Draft of chapter in book edited by Frederick G. Conrad and Michael F. Schober, entitled *Envisioning the Survey Interview of the Future*, published by Wiley and Sons, Inc.

Send correspondence to:

Art Graesser
Psychology Department
202 Psychology Building
University of Memphis
Memphis, TN, 38152-3230
901-678-2742
901-678-2579 (fax)
a-graesser@memphis.edu

There undeniably has been a revolution in new communication technologies during the last decade. Some of these advances in technology will end up having a radical impact on the process of collecting survey interview data, as the chapters in this edited volume make abundantly clear. A change in interview technology may end up being a necessity, not merely a convenience or a matter of cutting costs. For example, the survey world will need to make some adjustments very soon as the public experiences a shift from landline telephones to mobile telephones or to the web. Surveys will need to be very different in 10 years if the mainstream human-computer interface ends up having animated conversational agents, graphical user interfaces, speech recognition, and multimodal interaction. We can only speculate, of course, what the world will be like in 10 years.

This chapter has two primary goals. Our first goal is to summarize the existing state of research on interview technologies. We will not attempt to give an exhaustive and comprehensive summary, given the broad coverage of the field in the introductory chapter by Conrad and Schober (chapter 1). Instead, we provide a succinct snapshot of where we are now in the field in order to address important themes, obvious research gaps, and unresolved tradeoffs on the impact of technologies on psychological and practical dimensions. Our second goal is to identify some important research areas for the future.

We entirely recognize the need to be cautious when exploring the impact of technology on survey interview practice. Consider, for example, the role of technology in training and education. Cuban (1986, 2001) has documented that technology has historically had a negligible impact on improvements in education. As a case in point, visionaries predicted that radio and television would have a revolutionary impact on education, but this never materialized, as we all know. Clark (1983) argued that it is the pedagogy underlying a learning environment, not the technology *per se*, that typically explains learning gains. That conclusion of course motivates the need to investigate how particular technologies are naturally aligned with particular psychological principles, theories, models, hypotheses, or intuitions. These lessons learned from the learning sciences have relevance to the present book on interview technologies.

The *universal interface* of any communication system is the human face, with accompanying speech, gesture, facial expressions, posture, body movements, and human intelligence. This face-to-face (F2F) “inter - face” has sometimes been considered the gold standard for survey interviews. All other technologies are mediated, if not degenerate, forms of F2F communication, whether they be surveys with paper and pencil, telephone, web, or animated conversational agents. One fundamental question is what gets lost or gained from these alternatives to F2F interviews. For example, compared to web surveys, F2F has an advantage of being better able to detect whether the respondent understands the survey questions by virtue of paralinguistic cues, e.g., pauses, intonation, speech disfluencies, gaze, posture, and facial expressions. These cues provide a *social presence* that is minimal or nonexistent on web surveys. However, on the flip side, the social presence makes respondents more reluctant to share information on sensitive topics such as sexual activities and consumption of drugs and alcohol (Currivan, Nyman, Turner, & Biner, 2004; Tourangeau & Smith, 1996). There is a fundamental tradeoff between technologies that afford social presence and the revelation of sensitive information (Sudman & Bradburn, 1974). Such a comparison in technologies suggests that it is illusory to believe that there is a perfect universal interview technology. Instead, there are tradeoffs between evaluation measures when considering the landscape of technologies. We need to develop a science that takes these tradeoffs into account when selecting the optimal technology for a survey that is targeted for a particular topic, objective, population, budget, and margin of accuracy.

Current State of Research on Interview Technologies

Table 1 lists the interview technologies that have been used in the survey field or investigated by researchers of survey methodology. Chapters in this volume focus on particular technologies, as in the case of video conferencing (Anderson, chapter X), interactive voice response (Bloom, chapter, X; Couper, chapter X), mobile and landline telephone surveys (Fuchs, chapter X; Hancock, chapter X), instant messaging (Hancock, chapter X), web surveys (Couper, chapter X; Fuchs, chapter X), computer mediated conversation (Fussell et al., chapter X), multimodal interfaces (Johnston, chapter X), computer assisted interviews (Bloom, chapter X; Couper, chapter X), advanced sensing technologies (Bloom, chapter X; Cassell & Miller, chapter X; Johnston, chapter X), and animated conversational agents (Cassell & Miller, chapter X; Person, chapter X). Most of these chapters compare two or more interview technologies in an attempt to assess the impact of such contrasts on performance measures and also to test general psychological principles.

Table 2 lists the characteristics on which the interview technologies have been specified, measured, and evaluated. The table identifies the technology's affordances (Clark & Brennan, 1991, chapter by Fussell et al.), modalities and media (Anderson, chapter X; Couper, chapter X; Couper, Singer, & Tourangeau, 2004; Johnston, chapter X, Johnston & Bangalore, 2005; Mayer, 2005;), conversational capabilities (Conrad & Schober, 2000, chapter X; Schaeffer & Maynard, 2002, chapter X; Schober & Conrad, 1997, 2002; Suchman & Jordan, 1990), errors (Groves, 1989, chapter X), costs, and ethical ramifications (Konstan, chapter X; Marx, chapter X). Some of these characteristics have been investigated in considerable detail, such as validity and reliability, whereas data are conspicuously absent for costs of development and implementation.

*** INSERT TABLES 1 AND 2 ABOUT HERE ***

At this point in the science it would be worthwhile to fill out all of the cells in a matrix that we call the *technology-characteristic landscape* (TC-landscape). In essence, the affordances, modalities, media, conversational capabilities and other characteristics in Table 2 are specified in tables that map these characteristics onto particular interview technologies in Table 1. Researchers have filled out parts of the landscape. For example, Fussell et al. (chapter X) and Hancock (chapter x) identify the affordances for F2F, video conferencing, telephone, email, versus instant messaging. The conventional technologies (F2F, self administered questionnaires, telephone) have been compared on reliability and validity (Groves, 1989). However, cells are sparse on the more advanced interview technologies such as animated conversational agents and multimodal interfaces.

Regarding costs, it is informative that quantitative data are either missing or difficult to locate for the advanced interview technologies. The comparative costs can be estimated at a crude level. For example, development costs are comparatively high for developing IVR and web surveys compared with F2F and telephone surveys with humans, but the implementation costs would yield the opposite trend. Without precise costs on development and implementation, no solid conclusions can be made on return on investments (ROI's) for alternative interview technologies. An econometric analysis would be worthwhile that allows an assessment of the gains in validity or reliability as a function of increments in costs for particular technologies.

Psychological theories and principles offer predictions on the comparative impact of alternative technologies on measures of interest. *Social presence* (Reeves & Nass, 1996) is a theoretical principle that has received considerable attention. An interview technology has progressively more social presence to the extent that it resembles human characteristics of the gold standard F2F interview. This would predict the following gradient on social presence: F2F > telephone > IVR > self-administered questionnaire. As mentioned earlier, an interview technology with high social presence has the liability of lowering the likelihood that respondents will accurately report information on sexual habits, consumption of alcohol, and other sensitive topics. Precisely where animated conversational agents fit in this continuum is unclear compared to telephones because agents have a face, unlike the telephone interviews, but the speech and conversational intelligence is limited. Agents may be indistinguishable from F2F in highly scripted rigid interviews that have few degrees of freedom, but the agents would be akin to IVR when respondents desire a more flexible conversation that permits clarification questions. Conversational interviews require human intelligence that successfully achieves deep comprehension, language generation, and the processing of pragmatically appropriate paralinguistic cues, all of which are beyond the horizon of current agent technologies. Of course, this does not stop researchers from attempting to build agents that emulate humans. Some reasonably convincing agents have been developed in the context of tutoring and advanced learning environments, as in the case of AutoTutor (Graesser, Chipman, Haynes, & Olney, 2005; Graesser, Person, & Harter, 2001; VanLehn, Graesser, et al., in press) and Tactical Iraqi (Johnson & Beal, 2005).

Conversational grounding is another theoretical principle that has salient relevance to interviews (Conrad & Schober, 2000, chapter X; Schaeffer & Maynard, 2002, chapter X; Suchman & Jordan, 1990). Grounding is permitted when the respondent can ask questions or otherwise receive help that clarifies the intended meaning of terms and ideas expressed in survey questions. This is impossible on self-administered questionnaires but possible in F2F and telephone interviews that allow interviewers to answer respondents' clarification questions. It is also possible on web surveys with facilities that answer respondent questions (Schober, Conrad, Ehlen, & Fricker, 2003). Conversational interviewing increases the likelihood of there being an alignment between the plans, intentions, and common ground of interviewer and respondent. One danger in conversational interviewing, however, is that the interviewer runs the risk of implicitly unveiling attitudes, leading the respondent, and thereby increasing error. An agent would of course provide a more controlled conversational interaction, but this does not necessary reduce the error. A quirky agent would conceivably create pragmatic irregularities that also result in sources of error.

Discourse theories emphasize the importance of identifying the structure, ground rules, and constraints of specific conversational registers. Bradburn (chapter X) identified the four core characteristics of traditional survey interviews (see also Fowler & Mangione, 1990): (1) it is a special kind of conversation with its own rules, (2) there is a sequence of questions posed by the interviewer and answers provided by respondents, (3) the interviewer's questions control the conversation, and (4) there is a stereotypical question format. Conversational interviewing loosens the constraints by allowing the interviewer to answer respondents' clarification questions (a limited form of mixed initiative dialogue), and for the question formats to vary in an attempt to achieve conversational grounding and alignment. It is important for the researcher to specify the characteristics of the discourse register of any interview technology.

Contributors to this volume have meticulously identified many of the benefits and liabilities of particular interview technologies. It is quite apparent that there are tradeoffs when examining many of the characteristics and measures. There is a tradeoff between: social presence and the reporting of sensitive information, between the use of computer technologies and implementation costs, between complex interfaces with many options and the need for training, between speech and interpretation accuracy, between advanced technologies and errors of coverage and non-response -- the list goes on. It is important to document these tradeoffs as researchers fill in the cells of the TC landscape.

Survey researchers have developed coding systems to track both the verbal content of the answers and the unconscious reactions, paralinguistic communication cues, and other forms of paradata (Presser et al., 2004; Lessler & Forsyth, 1996). The verbal codes include categories of speech acts, such as read survey item, repeat survey item, request clarification, repeat answer, signal return to script, laugh, and so on (Schaeffer & Maynard, 2002, chapter X). The unconscious codes are pauses, restarts, speech disfluencies, intonation contours, eye gaze, facial expressions, gestures, posture, and so on (Cassell & Miller, chapter X). These unconscious codes, measured by trained judges or instruments, are sometimes diagnostic of respondents' lying (Hancock chapter x), but so are the verbal codes. One question for the Institutional Review Board (IRB) committees is whether informed consent is needed when researchers analyze the unconscious communication channels (Konstan, chapter X; Marx, chapter X). Informed consent is normally required for the verbal answers and actions that respondents intentionally provide, but not currently for the unconscious streams of activity.

Matters of ethics of course need to be scrutinized for all of the interview technologies. There are *seven critical issues* relevant to ethical violations in clinical and medical research according to Emmanuel, Wendler, and Grady (2000): (a) social or scientific value, (b) scientific validity, (c) fair subject selection, (d) favorable risk-benefit ratio, (e) independent review, (f) informed consent, and (g) respect for enrolled subjects. Fair subject selection is called into question when subsets of the target population tend to be excluded and thereby create coverage error or non-response error. If elderly and low SES populations tend not to use the web, then a web based survey would be prone to this ethical violation (Couper, 2000). Informed consent is not collected in web surveys that are routinely part of the practice of some marketing research firms and for the videotaped surveillance of citizens. The ubiquitous and noninvasive methods of data collection are not always submitted to IRB's so they are prone to ethical violations.

The current state of survey interview research is underdeveloped when it comes to examining differences in culture, languages, community norms, and demographic characteristics other than race, gender, and SES. Cross cultural research is conspicuously nonexistent in the more advanced interview technologies. The TC landscape would be tremendously expanded as we add these socio-cultural dimensions.

Research Needs in the Near Horizon

This section identifies five important directions for survey researchers to pursue in the future. Some of these would help fill gargantuan gaps in the TC landscape. Others would increase the quantitative and analytical sophistication of the research.

(1) *Fine-Grained Engineering and Quantitative Modeling*

There comes a point in a social science when researchers shift from inferential statistics that compare conditions on dependent measures to quantitative models that would appeal to engineers. It is time to wear some engineering hats as we move forward to the next stage of analyzing interview technologies. There are a suite of engineering models that have been used in the fields of human-computer interaction, cognitive engineering, and human factors.

One class of models is state-transition networks. There are a finite set of states and a set of arcs that designate transitions between states. Figure 1 shows a simple state transition network for survey interactions. There are five states that designate the speech acts of the interviewer and respondent (nodes A through E) and 8 arcs that designate transitions between the speech acts (arcs 1 through 8). All surveys require node A (interviewer asks survey question), transition 1, and node B (respondent answers question); this can be can be signified as either A → B, A1B or simply AB.

*** INSERT FIGURE 1 ABOUT HERE ***

Interview technologies vary on what speech acts are admissible or available in the interview. A self-administered questionnaire would afford only the AB path because there is no possibility of respondents asking clarification questions (node C), the interviewer answering these questions (node D), and the interviewer formulating revised questions (node E). The only option available to the survey researcher is to optimize the context and wording of the questions to maximize the reliability and validity of the responses. This is accomplished by pretesting the questions (Presser et al., 2004) or by using computer tools, such as Question Understanding Aid (QUAID, Graesser, Cai, Louwerse, & Daniels, 2006), to minimize the respondents' misunderstanding the questions.

A conversational interview, on the other hand, would allow speech acts in nodes C, D, and E in an effort to enhance conversational grounding and alignment (Schaeffer & Maynard, 2002, chapter X; Conrad & Schober, 2000, chapter X; Schober & Conrad, 1997; Suchman & Jordan, 1990). This is illustrated in the following hypothetical exchange.

Interviewer Question (A): *How many vehicles are in your household?*

Respondent Answer (B): *20 or 30*

Interviewer Revised Question (E,A): *How many vehicles during the last year?*

Respondent Question (C): *Do you include motorcycles and bicycles?*

Interviewer Answer (D): *Vehicles included motorcycles, but not bicycles.*

Respondent Answer (B): *Okay, 4.*

This example illustrates that the initial answer of *20 or 30* would not be valid, that the self-administered questionnaire would fall prey to the invalid response, and that the conversational interview would unveil a more valid response. The potential downside of the conversational interview is that the interviewer's speech acts run the risk of telegraphing expectations of the interviewer and indirect feedback, another source of error. So there is a trade-off between validity and these types of biases.

Ideally the survey methodologist would be able to collect quantitative parameters, metrics, and measures for the nodes and arcs in Figure 1. For each of the interview technologies, researchers would have an estimate that *any* answer is produced by the respondent (node B), that a *reliable* answers is produced, and that a *valid* answer is produced. Survey methodologists collect such data in their research, although it is widely acknowledged that estimates of validity are extremely difficult to collect in practice. What is being advocated is an even more refined analysis. We can quantify the likelihood of interviewers expressing speech acts {C | D | E}, the extent of different types of errors, correlations between different types of error and {C | D | E}, and transaction times for completing C, D, and E. With such quantitative indices at hand, we can simulate survey completion times and the extent that errors will be committed with a new technology before it is designed. For example, one could imagine a web survey that answered respondent's clarification questions through a word definition help facility (allowing C and D), but would not be able to formulate revised questions (node E). Performance measures could be simulated for such a system. The designer could also change the interface to encourage the respondent to ask clarification questions (node C) by instructions or highlighting the help facility. Designers could then explore how that interface enhancement would influence survey completion times, the validity of answers, and other performance measures. The state transition network and affiliated parameters would generate estimates in its simulations. The model would be validated by comparing the predicted quantities with empirical data that is eventually collected.

Another example of a quantitative model is the *GOMS* model of human-computer interaction (Card, Moran, & Newell, 1983; Gray, John, & Atwood (1993). GOMS stands for Goals, Operators, Methods, and Selection Rules. This model was originally used to simulate performance of users of word processing software but has since than been applied to dozens of tasks with system-person interfaces, including telephone operator systems. Performance is simulated at multiple levels of grain size: the keystroke level, completion of subgoals, and completion of major tasks. Another virtue of GOMS is that it uses a theoretical model of human information processing to predict execution times for basic operations of perception, action, and memory. It would predict, for example, how long it takes to: (a) to process one eye fixation in a perception-cognition-motor cycle (approximately 240 milliseconds), (b) move the eyes from one location on a computer screen to another location, (c) move a finger from one location to a target location of a particular size, and (d) make a decision among N alternatives.

One could imagine the benefits of such a model in the arena of animated conversational agents. The GOMS model would simulate how long it takes for an agent to produce a spoken answer to a respondent’s question versus how long the same answer would take for the respondent to read. Given that a spoken message is difficult to ignore, and that the respondent would sometimes skip reading a printed message, the researcher can estimate the comparative advantage of the two media when measuring the likelihood that the respondent attends to the message. Many other questions could be answered by the modeling tool. How would the likelihood of attending to the message be influenced by the placement of the printed message on a screen? How would listening times and the respondents’ requests for repeated messages be influenced by agents speaking at different speeds, with different dialects, and different paralinguistic channels. Answers to such questions, and literally thousands of other questions, can be estimated with a GOMS model. It is important to acknowledge that there is never enough time for survey methodologists to test empirically a staggering number of alternative interfaces. We would argue, therefore, that such a quantitative model is a necessity – not merely a luxury -- for survey methodologists during this age of rapid changes in technologies.

(2) ***Analysis of Dialogue Modules***

There is a need to perform a deep analysis of the discourse that underlies an interview technology because all surveys and human-computer interfaces are fundamentally forms of discourse. It is convenient to decompose the discourse into dialogue modules. A dialogue module is defined, for the present purposes, as a dialogue exchange between the interviewer and respondent that accomplishes a particular discourse function. The dialogue moves in Figure 1, for example, would include the following dialogue modules: (a) Respondent answer survey question asked by interviewer, (b) interviewer clarify meaning of expression for respondent, and (c) interviewer align question with knowledge of respondent. Each module has its own participant roles, knowledge and beliefs of participants, goals or intentions of participants, and pragmatic ground rules. We believe that it is important for survey researchers to understand the cognitive, pragmatic, and social components of the important dialogue modules in interviews and to systematically analyze how these components are realized (if at all) in each interview technology under consideration.

Consider the main dialogue module below that would apply to all surveys and interview technologies.

Dialogue Module 1: *Respondent answer survey question asked by interviewer.*

1. I = interviewer; R = respondent; Q = question

2. I does not know answer to Q at time t_1

3. I believes that R knows the answer to Q at time t_1

4. I believes that R understands the meaning of the Q at time t_1

5. I asks R the Q at time t_1

6. R understands the meaning of the Q at time t_2

7. R knows the answer to the Q at time t_2

8. R answers the Q at t_2

9. R believes I understands R’s answer to the Q at time t_3

10. I believes that R understands the meaning of the Q at time t_3

11. I understands the meaning of the answer at time t_3

12. I believes that R believes that the answer to the Q is true at time t_3

13. I acknowledges R’s answer to the Q at time t_4

14. I accepts R’s answer to the Q at time t_4

15. R believes that I accepts R’s answer to the Q at time t_5

At first blush this may seem to be a tedious specification of obvious epistemological and pragmatic assumptions of a simple survey question and its answer. However, each of these (as well as others if we were to include intentions) has important ramifications on interview technologies and the success of a conversational exchange. Fortunately, computational linguists have developed computer tools to assist researchers in keeping track of the beliefs, knowledge plans, intentions, and other states of participants in dialogue modules. Two of these are COLLAGEN (Rich, Sidner, & Lesh, 2001) and the TRINDI toolkit (Larsson & Traum, 2000).

The 15 components in the example dialogue module would apply to a successful exchange with satisfactory alignment between the interviewer and respondent. However, the verbal answers and paradata in a F2F conversational interview often suggest that one or more of the assumptions are false. If the respondent pauses, frowns, or looks confused at time t_2, then assumptions 6 and 10 are suspect; this would motivate the interviewer to revise the question in a conversational interview and launch a subordinate dialogue module: interviewer align question with knowledge of respondent. If the interviewer pauses, has a hesitation prosody in the acknowledgement, looks skeptical, or asks a revised question at time t_4, then assumptions 6, 9, 10, 14, and/or 15 are suspect; this would motivate the respondent to ask a clarification question and the following subordinate dialogue module would be launched: interviewer clarify meaning of expression for respondent. The subtleties of the paradata are often diagnostic of which assumptions are suspect.

Communication breaks down when there is misalignment at different levels of communication (Clark, 1996; Pickering & Garrod, 2004; Walker et al., 2003). Misalignments would of course compromise the validity of the responses in surveys, as has been shown in the research by Conrad and Schober (Conrad & Schober, 2000; Schober & Clark, 1997). In the example dialogue module, there can be misalignments in components 4-6-10, 9-11, 7-14, and so on. Potential misalignments become visible in some interview technologies more than others. F2F and video conferencing allow visual and auditory paradata from the respondent, telephone allows only auditory, whereas self administered questionnaires and the web allow neither. Limitations in paradata of course have an impact on errors and other measures. Designers of computer technologies might provide special facilities to compensate for the lack of paradata in those technologies without contact with a human interviewer. For example, a web survey might have a help facility that answers the respondents' clarification questions (nodes C and D in Figure 1) or that generates revised questions (node E). However, it is well documented that individuals rarely ask questions and seek electronic help in most computer systems (Carroll, 1987) and learning environments (Graesser, McNamara, & VanLehn, 2005). The likelihood of help seeking is no doubt lower in a survey environment because the respondent is voluntarily *providing* information rather than having the goal of *obtaining* information (Schober et al., 2003). It is not sufficient for a computer technology to simply offer electronic assistance for conversational grounding and alignment because there is a low likelihood that the user will use the facilities when they are needed. Conversational agents hold some promise in encouraging the respondent to seek help or in barging in to offer help. However, there is a risk of the agents committing false alarms when the agent cannot accurately understand the respondent and diagnose misalignments.

(3) ***Measuring the Accuracy of Interpretation and Sensing Components***

How accurately does the interviewer, computer, or instrument measure the respondents' intended messages and unintended paradata? This is a pervasive research question in survey methodology, and indeed the field of communications in general. Speech recognition has received considerable attention because this is a technology that would have enormous practical value in the world of surveys (Bloom, chapter X; Johnston, chapter X). Although the quality of speech recognition is improving (Cole et al., 2003), the accuracy of speech-to-text translation is substantially higher for humans than computers (Meyer et al., 2003).

The accuracy of spoken messages being recognized correctly apparently improves quite a bit when the speech can be accompanied by pointing to interface elements via touch panel or electronic pens (Johnston, chapter X). One example of a successful multimodal interface is the MATCH (Multiple Access to City Help) system (Johnston, chapter x). More research is needed on how well respondent messages can be automatically interpreted in a multimodal system with various combinations of speech recognition, handwriting recognition, gesture recognition, recognition of facial expressions, detection of eye gaze directions, detection of posture, and so on. In addition to the accuracy of the verbal messages, there needs to be research on the accuracy of sensing the paradata, such as pauses, intonation, direction of gazes, gesture, posture, and facial expressions. In our own lab, we have attempted to classify learners’ emotions while college students interact with AutoTutor (D’Mello, Craig, & Graesser, in press). Emotions such as confusion, frustration, boredom, engagement, delight, and surprise are being classified on the basis of the verbal dialogue, facial expressions, speech intonation, and posture. Our hypothesis is that emotion classification will require a combination of communication modalities, not any one alone.

The messages and paradata of the respondent are of course influenced by those of the interviewer. Such interviewer-respondent interactions need to be incorporated in any assessment of the accuracy of interpreting respondent messages and paradata. Respondents’ conversational style and paradata can model, mirror, or otherwise respond to what the interviewer does. One of the virtues of conversational agents is that the messages and paralinguistic cues are controlled whereas those of human interviewers are variable or intractable. Future research with agents in interviews need to document the impact of the agent’s conversational, physical, stylistic, and personality characteristics on the behavior of the respondents (see Cassell & Miller, chapter X; Person, chapter X).

The accuracy of automated comprehension of messages is limited at this point in the fields of computational linguistics and artificial intelligence (Allen, 1995; Jurafsky & Martin, 2000; Rus, 2004). There have been impressive advances in computer analyses of words (such as *Linguistic Inquiry Word Count* , Pennebaker & Francis, 1999), shallow semantics (such as *Coh-Metrix*, Graesser, McNamara, Louwerse, & Cai, 2004), automated grading of essays (such as *e-Rater*, Burstein, 2003; *Intelligent Essay Assessor*, Landauer, Laham, & Foltz, 2000), and question answering (Dumais, 2003; Harabagiu, Maiorano, & Pasca, 2002; Voorhees, 2001). However the performance of automated systems is far from adequate whenever there is a need for the construction of context-specific mental models, knowledge-based inferences, deictic references, resolution of anaphoric references, and other components of deep comprehension. Surveys require a high precision in comprehending messages, unlike other conversational registers that can get by with less rigorous grounding and alignment of meanings, such as tutoring (Graesser et al., 2001, 2005) and small talk (Bickmore & Cassell, 1999; Cassell et al., 2000). An interviewer agent would require a fairly accurate mental model of the respondent (Walker et al., 2003), which is at a least a decade away from being technically achieved.

On a technical note, the various fields that have investigated communication systems have adopted somewhat different quantitative foundations for measuring communication accuracy. Social psychologists are prone to collecting ratings from multiple judges as to whether a message is understood by a respondent or system. Computational linguists collect recall scores, precision scores, and F-measures in comparisons between a computer's output and judgments of human experts (Jurafsky & Martin, 2000). Cognitive psychologists collect hit rates, false alarm rates, d' scores, and other measures from signal detection theory in comparisons between computer output and judgments of humans. It would be worthwhile for colleagues in the survey methodology world to converge on some agreement on measurements of the accuracy when analyzing interpretation and sensing components.

(4) *Animated Conversational Agents in the Survey World*

Our forecast is that animated conversational agents will eventually become ubiquitous in human computer interfaces, in spite of the cautions that problems will arise when humans attribute too much intelligence to the agents (Norman, 1994; Shneiderman & Plaisant, 2005). Adults have the ability to gauge the intelligence, personality, and limitations of other humans and classes of humans. Adults can discriminate what is true in the real world from what they see in movies. In essence, they can make fine discriminations about the capabilities humans and their presence in technologies. This being the case, we can imagine that there will be a new genre of agents that are *Auto-Interviewers* and that these agents will be eventually understood and accepted by the public. However, it us uncertain how the public will view the Auto-Interviewer's intelligence, believability, trust, pragmatic ground-rules, personality, and other dimensions of social presence (Cassell & Miller, chapter X; Cassell, Sullivan, Prevost, & Churchill, 2000; Reeves & Nass, 1996). An agent can announce the limits of its protocol ("Sorry I am not allowed to answer your questions") or capabilities ("Sorry but I am not understanding you", "Could you speak more clearly?"), just as interviewers and other people do. We could imagine an Auto-Interviewer with an engaging and endearing personality that respondents enjoy interacting with for a sustained period of time, despite its limitations on various human dimensions. We could imagine a new class of agents with its own identity: the Auto-Interviewers.

It will be important to conduct parametric studies that investigate the impact of agent features on respondents. Is it best to have agents matched to the respondent in gender, age, ethnicity, and personality, or is it better have contrasts (e.g., male agents with female respondents) or high status prototypes (earnest older white males)? How do these features and the agent's attractiveness influence the engagement, completion rates, and response validity of respondents? How and when should the agent give back-channel feedback (uh huh, okay, head nod) that the agent has heard what the respondent has said? Should the agent deviate from realism, as with cartoon agents or caricatures, or be very close to a realistic depiction of humans?

The roboticist Masahiro Mori (1970) claimed that people have an increasing level of comfort and even affection for robots as they increasingly resembles a human. However, there reaches a point on the robot-human similarity continuum that the human feels uneasy or even disgusted when the robot mimics a human very closely, but not quite perfectly. Humans become more comfortable again when there is a perfect robot-human match, a state that is theoretical because no robots are that good. Mori calls the uneasy trough the *zone of the uncanny valley*. The technology is now available to measure human's reactions to agents as a function of the different values on the agent-human similarity continuum (Conrad & Schober, chapter X). There is software developed by Neven Vision (recently acquired by Google) that has agents mirror what a human agent does by mapping the facial features of the agent onto the features of the human. This will allow researchers to empirically test the uncanny valley and other mappings between agent characteristics and human emotions.

A successful animated agent is not limited to its appearance. It will need to coordinate the speech, facial expressions, gesture, eye gaze, body posture and other paralinguistic channels. The conversational dialogue will need to generate the appropriate speech acts at each turn. A totally natural agent would need to generate speech disfluencies that reflect the difficulty of its message planning mechanism and its conversational floor management (i.e., maintaining its turn or taking the conversational floor). There is a rich terrain of research avenues for those interested in animated conversational agents.

Once again, skeptics may ask "why bother.?" Our view is that a good Auto-Interviewer will represent the human-computer interface of the future, will be cheaper to implement (although expensive to initially develop), and will provide more control over the interview than will F2F and video conference interviews with humans.

(5) ***Culture***

It is perfectly obvious that our research on interview technologies needs to be replicated in different cultures in order to assess the scope and generality of scientific claims. This gives all of us free reigns to tour the world in the interest of multi-cultural research on survey methodology. We can explore different cultures, languages, age groups, socioeconomic strata, and so on.

The role of technology in different cultures is particularly diverse. The younger generations tend to use advanced technologies, but not the elderly, so there is the worry of coverage and non-response errors in the elderly populations (Groves, chapter X). People in most cultures do not understand the semiotics of visual icons that bloat the interfaces of most web sites, so they tend to be excluded from web interviews and many other interview technologies. The study of icon semiotics is indeed an important area for exploration because the computer interfaces of today are too arcane and unintuitive for the majority of the US population, a culture that values technology. Designers often justify their poor human-computer interfaces by claiming that users can read instructions or learn from their peers. However, the data reveal that

US adults quickly abandon any web site or inexpensive software product if they encounter a couple of obstacles within a few seconds of receiving it (Zachary, 200X). The truth is that the open market of software users is unforgiving of shabby human-computer interfaces.

The face is the universal interface, as we proclaimed at the beginning of this chapter. As such, the face presents very few obstacles in human-computer interaction for members of different cultures. Those of us who conduct research with animated conversational agents know that the time it takes to instruct participants how to use a computer system with an agent is on the order of a few seconds, with few if any user errors. In contrast, it takes several minutes to instruct participants how to read and use a conventional interface, often with troublesome bottlenecks that require retraining. As expressed earlier, however, there are potential liabilities of these agents (e.g., technical imperfections, social presence lowering rates of sensitive disclosure) and these probably get magnified when considering differences among cultures. However, this latter claim about differences among cultures is speculative and awaits empirical research.

The younger generation in the US is currently a generation that thrives on computer games (Gee, 2003) and instant messaging (Kinzie, Whitaker, & Hofer, 2005). The older generation is virtually alienated from these two technologies. What are the implications of this fact? Should we be embed our surveys in games and instant messaging if we want to reach the next generation? The younger generation is also in a culture of media that cultivates short attention spans. Should we distribute the survey questions over time, locations, and contexts in order to improve response rates? Should the surveys fit within these constraints?

Citizens in the workforce live in different cultures than the youth and the elderly. Those in the workforce have very little time and that may prevent some from being willing to complete a survey. Workforce respondents may be more open to completing a survey if they are viewed as consultants or experts on a topic. The solution to reducing non-response rate may lie in creating more engaging interviews that are perceived as being important to members of a culture.

Closing Comments

This chapter has provided a snapshot of the current state of research in learning technologies and has proposed some research directions that are needed in the near horizon. This is an unusual point in history because we are in a revolution of communications technologies and the communication media of choice are extremely diverse among different sectors of the population. Mail is no longer reliable, nor is the telephone, nor the web. The media of choice are entirely different for the young and old, for the rich and poor, and for those in different languages and cultures. The future of survey research is destined to be entirely different than it was 30 years ago. We have no choice but to pursue novel solutions and to shed some of the sacred guidelines from the past.

Acknowledgments

This research was partially supported by the National Science Foundation (ITR 0325428, REESE XXXXX). Any opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF.

References

Allen, J. (1995). *Natural language understanding*. Redwood City, CA: Benjamin/Cummings.

Bickmore, T., & Cassell, J. (1999). Small talk and conversational storytelling in embodied conversational characters. *Proceedings of American Association for Artificial Intelligence Fall Symposium on Narrative Intelligence* (pp. 87-92). Cape Cod, MA: AAAI Press

Burstein, J. (2003). The E-rater scoring engine: Automated essay scoring with natural language processing. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 133-122). Mahwah, NJ: Erlbaum.

Card, S., Moran, T., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Erlbaum.

Carroll, J.M. (1987)(Ed.) *Interfacing thought: Cognitive aspects of human-computer interaction.* Cambridge: MIT Press/Bradford Books.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (2000). *Embodied conversational agents*. Cambridge: MIT Press

Clark, H.H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*, 445-460.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. Resnick, J. Levine, & S. Teasely (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, D.C.:American Psychological Association.

Cole, R. van Vuuren, S., Pellom, B., Hacioglu, K., Ma, J., Movellan, J., Schwartz, S., Wade-Stein, D. Ward, W., & Yan, J. (2003). Perceptive animated interfaces: First steps toward a new paradigm for human computer interaction. *Proceedings of the IEEE, 91*, 1391-1405.

Conrad, F.G., & Schober, M.F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly, 64,* 1–28.

Couper, M.P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly, 64*, 464-494.

Couper, M.P., Singer, E., & Tourangeau, R. (2004). Does voice matter? An interactive voice response (IVR) experiment. *Journal of Official Statistics, 20 (3),* 1-20.

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920*. New York: Teachers College.

Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Cambridge, MA: Harvard University Press.

Currivan, D., Nyman, A.L., Turner, C.F., & Biener, L. (2004). Does telephone audio computer-assisted survey interviewing improve the accuracy of prevalence estimates of youth smoking? Evidence from the UMass Tobacco Study. *Public Opinion Quarterly, 68,* 542-564.

Dillenbourg, P., & Traum, D. (2006). Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *Journal of the Learning Sciences, 15*(1), 121-151.

D'Mello, S.K., Craig, S.D., & Graesser, A.C. (in press). Predicting affective states through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education, 16,* 3-28.

Dumais, S. (2003). Data-driven approaches to information access. *Cognitive Science, 27*(3), 491-524.

Emmanuel, E. J., Wendler, D., & Grady, C. (2000). What makes clinical research ethical? *Journal of the American Medical Association, 283,* 2701-2711.

Fowler, F.J., & Mangione, T.W. (1990). *Standardized survey interviewing: Minimizing interview-related error*. Newbury Park: Sage.

Gee, J. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.

Graesser, A.C., Cai, Z., Louwerse, M., 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.

Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions in Education, 48,* 612-618.

Graesser, A.C., McNamara, D.S., Louwerse, M.M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers, 36,* 193-202.

Graesser, A.C., McNamara, D.S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist, 40,* 225-234.

Graesser, A.C., Person, N., Harter, D., & TRG (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education, 12,* 257-279.

Groves, R. M. (1989). *Survey errors and Survey costs*. New York: Wiley.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world performance. *Human-Computer Interaction, 8*(3), 237-309.

Harabagiu, S. M., Maiorano, S. J., & Pasca, M. A. (2002). Open-domain question answering techniques. *Natural Language Engineering, 1,* 1-38.

Johnson, W.L., & Beal, C. (2005). Iterative evaluation of a large-scale intelligent game for language learning. In C. Looi, G. McCalla, B. Bredeweg, and J. Breuker (Eds.), *Artificial Intelligence in Education: Supporting learning through intelligent and socially informed technology* (pp. 290-297). Amsterdam: IOS Press.

Johnston, M., & Bangalore, S. (2005). Finite-state multimodal integration and understanding. *Journal of Natural Language Engineering, 11,* 159-187.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice-Hall.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13,* 351-371.

Kinzie, M. B., Whitaker, S. D., & Hofer, M. J. (2005). Instructional uses of instant messaging (IM) during classroom lectures. *Educational Technology and Society, 8*(2), 150-160.

Landauer, T.K., Laham, D., & Foltz, P.W. (2000). The Intelligent Essay Assessor. *IEEE Intelligent Systems 15* , 27-31.

Larsson, S. & Traum, D. (2000). Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering***,** *6* (3-4), 323-340.

Lessler, J.T., & Forsyth, B.H. (1996). A coding system for appraising questionnaires. In N. Schwartz and S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 259-292). San Francisco: Jossey-Bass.

Mayer, R.E. (2005). *Multimedia Learning*. Cambridge, MA: Cambridge University Press.

Mori, M.(1970). Bukimi no tani (the uncanny valley). *Energy, 7,* 33–35. (In Japanese)

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

Norman, D. A. (1994). How might people interact with agents? *Communication of the ACM, 37*(7), 68-71.

Pennebaker, J.W., & Francis, M.E. (1999). *Linguistic inquiry and word count (LIWC).* Mahwah, NJ: Erlbaum.

Pickering, M.J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Brain and Behavioral Sciences, 27*, 169-190.

Presser, S., Couper, M.P., Lessler, J.T., Martin, E., Martin, J., Rothgeb, J.M., & Singer, E. (2004). Methods for testing and evaluating survey testing. *Public Opinion Quarterly, 68,* 109-130.

Reeves, B., & Nass, C. (1996). *The media equation.* New York: Cambridge University Press.

Rich, C., Sidner, C. L., & Lesh, N. (2001). COLLAGEN: Applying collaborative discourse theory to human-computer interaction. *AI Magazine, 22*(4), 15-25.

Rus, V. (2004). A first exercise for evaluating logic form identification systems, *Proceedings Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (SENSEVAL-3),* at the Association of Computational Linguistics Annual Meeting, July 2004. Barcelona, Spain: ACL.

Schaeffer, N.C., & Maynard, D.W. (2002). Occasions for intervention: Interactional resources for comprehension in standardized survey interviews. In D.W. 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. 261-280). New York: Wiley.

Schober, M. F., & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly, 60*, 576-602.

Schober, M.F., & Conrad, F.G. (2002). A collaborative view of standardized survey interviews. In D. Maynard, H. Houtkoop-Steenstra, N.C. Schaeffer, & J. Van der Zouwen (Eds.), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 67–94). New York: John Wiley & Sons.

Schober, M.F., Conrad, F.G., Ehlen, P., & Fricker, S.S. (2003). How web surveys differ from other kinds of user interfaces. In *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Shneiderman, B., & Plaisant, C. (2005). *Designing the user interface: Strategies for effective human-computer interaction* (Ed. 4). Reading, MA: Addison-Wesley.

Suchman, L. & Jordan, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association 85(409)*, 232-241.

Sudman, S., & Bradburn, N.M. (1974). *Response effects in surveys*. Chicago: Aldine.

Tourangeau, R., & Smith, T. W. (1998). Collecting sensitive information with different modes of data collection. In M. P. Couper, R. P. Baker, J. Bethlehem, J. Martin, W. L. Nicholls II, & J. M. O'Reilly (Eds.), *Computer Assisted Survey Information Collection* (431-453) . New York: John Wiley & Sons.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (in press). When are tutorial dialogues more effective than reading? *Cognitive Science*.

Voorhees, E. (2001). The TREC question answering track. *Natural Language Engineering, 7*, 361-378.

Walker, M., Whittaker, S., Stent, A., Maloor, P., Moore, J., Johnson, M., & Vasireddy, G. (2003). Generation and evaluation of user tailored responses in multimodal dialogue. *Cognitive Science*, *28*, 811-840.

Table 1. *Interview Technologies*

1. Face to face (F2F)
2. Video conferencing
3. Animated conversational agents
4. Video telephony
5. Telephone, including landlines and mobile phones
6. Interactive voice response (IVR)
7. Self-administered questionnaire via mail or directly administered
8. email
9. Instant messaging
10. Web
11. Computer assisted survey interview (CASI), including audio CASI and video CASI
12. Computer assisted personal interview (CAPI)
13. Computer assisted telephone interview (CATI)
14. Advanced sensing capabilities, including speech, gesture, and handwriting recognition
15. Multimedia and multimodal interfaces

Table 2. *Technology affordances, modalities, media, conversation capabilities, errors, costs, and ethics*.

Affordances (Clark & Brennan, 1991): Co-presence, visibility, audibility, cotemporality, simultaneity, sequentiality, reviewability, revisability

Modalities/media (Mayer, 2005): Visual, graphical, spatial, linguistic, print, auditory, spoken.

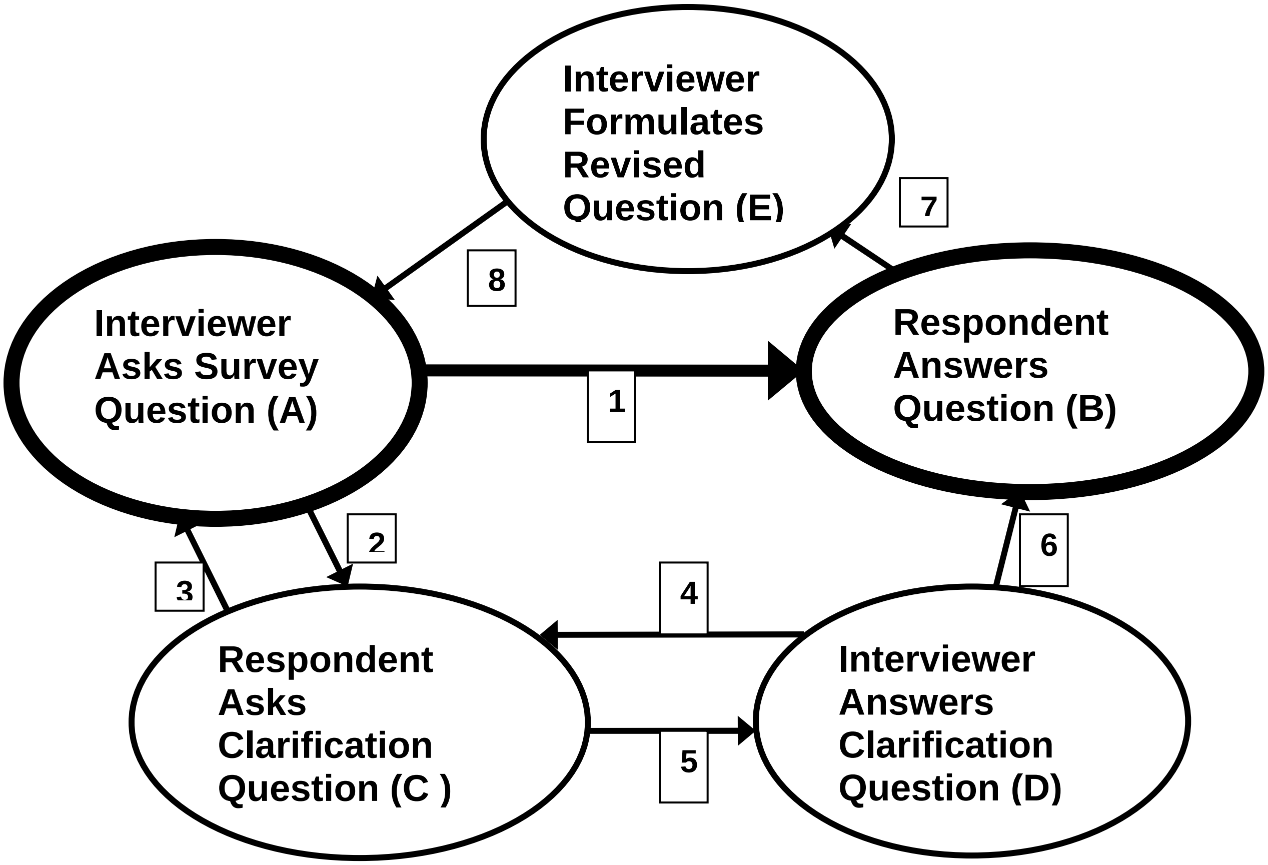
Conversation capabilities (Conrad & Schober, 1997; Schaeffer & Maynard, 19xx; chapters X and Y)**:** Social presence, scripted, interactive, grounded, interviewer-entered, mixed-initiative.

Errors (Grove, 199x, chapter X): validity, reliability, measurement error, processing error, coverage error, sampling error, non-response error

Costs: Development, implementation

Ethics. Privacy, informed consent, ethical principles

Figure 1: *A Simple State-Transition Network for Survey Interactions*.


Interviewer Formulates Revised Question (E)
7
8
Interviewer Asks Survey Question (A)
1
Respondent Answers Question (B)
2
3
4
6
Respondent Asks Clarification Question (C)
5
Interviewer Answers Clarification Question (D)