Jon Krosnick: So, Bob Belli is one of the leaders in the work that’s gone on in supplementing, kind of, ordinary interviewing with additional techniques, complex techniques for measuring events. And so, he is going to walk us through that work.

Robert Belli: Thanks, Jon, and thanks for the invitation to provide you with some of the kinds of things that I’ve been doing over the past several years; things that have been, if you will, governing my life and an opportunity to inform NSF in directions that I think this type of work should be moving in the future.

The name is Skip, is that right, Skip Lupia – talked about issues of legitimacy and what I really – when I, sort of, heard that and looked at what I want to talk about with you first, the issue of whether – what type of interviewing methodology one uses is an issue of legitimacy. And that is if one is interested in collecting retrospective reports, which is based on people, as best as we can, accurately reporting on their past, what kind of methodology, what kinds of methods, best promotes accuracy in reports?

And if you look at the goals of conventional standardized interviewing, it really is set up as a system by which one minimizes interviewer effects; that is, the notion is with the standard stimulus, all the variants that you get will be the variants in the respondent reports and not through the interviewers themselves, although that’s an ideal, it’s a goal, it’s a worthwhile goal in terms of trying to minimize interviewer effects. But, it’s main purpose is not to optimize the quality of respondents recording on their autobiographical past. And, hence, there may be a better strategy that we can use. And if one looks at the structure of autobiographical memory, there are various cues in that structure that will, should, assist people to remember, more fully, their past. And the strategies are based on a very simple principle. One of the best ways to get people to report more about their past more accurately in terms of reconstructing their past more accurately is to use whatever they have told you about, what they already can remember, as cues to a system to remember more. And because any of those kinds of cues are idiosyncratic to individuals, that is, what I have as a memory is going to differ from what you have as a memory and if those memories serve as adequate cues, then standardization becomes a real impediment in terms of providing standard stimuli because you can’t do that. Everyone’s situations are fairly unique.

There’s an also, perhaps, added benefit that, sort of, speaks to issues about conversation that we know a little bit about and not,
perhaps, as much as we would like to; and that is, there are aspects of conversation that help clarify meanings and clarify intent and there has been a literature out there that talks about standardization as actually providing some certain levels of impediments to the ordinary benefits of conversation.

Both calendar and time diaries, in many ways, do allow a method by which flexible interviewing, which is needed to maximize more accurate recall of one’s past, can occur. And they’re quite different instruments. The calendar instrument can go seek retrospective reports for decades, for the past year, for one’s entire life course. As Art Stone indicated earlier, time diaries are essentially a kind of method in which one is seeking information on what happened yesterday.

This is a calendar, which is implemented to gather information on unpleasant things and those unpleasant things have to do with exposure to intimate partner or domestic violence; such things as whether or not an intimate partner threatened to hit you, threw anything at you, punished, grabbed or shoved you, slapped you and so forth, many, many different items. In terms of how calendars work, what you can see is that there are various different domains of interest, such as places where people live, residence, where they went to school, places of work. And the thinking here and also, a very rich relationship history, the thinking in terms of cues is that initial states or initial partners one has had will help or assist people remember partners that they had later. And they may also help them remember partners they may have had earlier. I referred to that as sequential retrieval. It’s sequential retrieval process.

There’s also opportunities for parallel retrieval. That is in different domains, having lived at a certain place, could provide a good memory cue as to who your – or who was your relationship partner, your partner at a point in time.

And then, there’s top down retrieval, moving from more general to more specific; moving from the name of your relationship, name of your intimate partner to whether or not that partner had engaged in any of these domestic violence incidents.

As we move, this is a paper – this prior key owner was paper and pencil and it was a situation of face-to-face interviewing. As – when I was involved in the PSID, one of the issues there is that it’s a telephone interview. And the other issue is that we’re moving more towards computerized interviewing, CATI interviewing in this case. This was a calendar that was developed for purposes of
collecting information over the entire life course. It was a methodological study and it compared a calendar against a conventional question and I’ll talk about that in a minute. But, pretty much, the same principle hold here; that is, we collected – whoops, wrong – we collected earlier, if you will, information; not earlier in terms of the entire life course, but earlier during the interview, information on collecting a residential history. We then went into a employment history, these various tabs; went to the data entry sections of that interview and could use sequential retrieval in terms of remembering who the employers were at various points in time and cross reference that or use a parallel retrieval to point back to people’s residences.

But, in terms of looking at the basic ideas that calendar interviewing will produce better data quality in terms of retrospective reports than conventional, standardized questionnaires, there have been a number of studies that have actually done comparisons between the two. This is, if you will, just a brief – I just want to go over this fairly quickly in terms of just pointing out, in some of these studies, the calendar was used as an aid. Colleagues in the Netherlands, Wander van der Vaart and Tina Glassner has led this work. You can notice with regard to this research that they had faced various modes, face to face, paper and pencil, CATI instruments, web based instruments. We’ll talk about that more. And in this comparison, they’re using the calendars in aid to a conventional questionnaire, conventional alone and what they observed when they had validation data or going under the rubric that more reports is better, assuming that underreporting is the norm under these situations, which observes that the calendar plus conventional condition had done better than a conventional condition all of itself.

If we look to other work that has looked at calendars, invested alone versus conventional alone, we still observe, if you will, that calendars do better; not always, but more often in terms of accuracy report with different kinds of criteria. Here, more is better criteria; here, we have a validation set that consisted of PSID data that was asked in earlier interviews. With regard to that calendar that I showed you on intimate partner violence, I want to talk about a more reasonable pattern of results here. We’ll know things have been observed in these kinds of studies and other studies using conventional questionnaires is that there’s an age cohort effect in the sense that older persons will report exposure at later ages than younger persons. The age cohort effect, apparently, was due to recall problems because with the calendar, the age cohort effect was eliminated. And, again, validation studies, you
can also see various modes, telephone, face to face and computerized instruments as well.

Moving on, then, to the notion of a time diary. One of the things in which both of these methods, calendars and time diaries are similar, is the extent to which they’re based on people remembering temporal information and, hence, we have, with regard to the PSID child development supplement, they have a time diary that’s asking parents to report on their children what they did during the day – time beginning, time ending. And people then put sequences of different episodes. They also have what’s referred to as secondary activities. Moving forward and backward in time is a kind of sequential retrieval. Secondary activities, somewhat to parallel cueing, parallel retrieval. Art Stone already showed this, so I don’t have to go over this in too much detail, the American Time Use Survey. A very similar idea, but now, of course, it’s computerized. It’s a – this is a CATI instrument; whereas, if you looked at the PSID, this was self-administered. So, now we have CATI instrument. And some of the differences, which I think are worthwhile to point out, the ATUS asks people, provides both durations and beginning and stop times. So, people can either provide information on when an event stopped or how long it took. That duration versus timing, if you will, may be important. And if you look at the PSDI instrument, it only allows response report in terms of time, not in terms of duration.

There’s also, if you will, these precoded activities that interviewers can introduce. And, also, there then would be verbatims; that is, information about activities that are not precoded.

Some of the validity with the time drivers, I think, Art still went over this a bit more fully than what I’m going to do here. But, alternate sources of data, such as experience sampling, beepers, fairly high correlations, looking at spouse’s reports sometime, fairly high correlations; looking at an alternative validate – a source of validation that is actually looking at electrical output that was used and then looking at time and day patterns of energy use, high correlations and this notion of reduction in suspected over reporting, if you will, from standard question form estimates has been observed in time diaries as well. And Stan, you may want to, if you will, augment this at some point.

What is needed in terms of future work and some of this is already on going; some of it is not ongoing. One of the big areas, which needs a lot more concentration are areas with regard to visual design. Issues with regard to visual design, which has been
explored much more fully with more conventional, if you will, ways of asking questions and that, in terms of these various kinds of instruments and how they look and how they appear and how useable they are, really have not been as fully explored as they should be.

We have to get a much better understanding with regard to the interviewer-respondent interactional processes and how it is that language and memory interact. And one way to, sort of, gain some insight on that, which I’m going to show you, is engaging in more full behavior coding studies. There’s also now an emerging availability of paradata, which can, for example, provide information of the extent to which interviewers are interacting or interfacing with the instruments. And we can also begin to look at those kind of interfaces and data quality, just as I have here with behavior coding as well. And we also have to look at mode issues. I’m really fairly convinced – I think most of us probably would agree with this – sentiment; that self-administered, web based instruments is going to be the wave of the future. And, hence, we have to begin the process of actually understanding more fully how we can implement, if you will, these kinds of benefits of cues in a self-administered instrument and that is, in some ways, perhaps, replaced interviewers and calendars with an interviewer that may be more virtual and I’m going to talk about that in more detail as well.

Let me hint now or talk about real behavior coding and what’s involved in real behavior coding. It has, certainly, a range of a historical process in terms of a history of looking at interviewer performance, identify problematic questions. My interests are ones in which you want to examine the quality of verbal exchange between interviewers and respondents. There is a theory that I provided to you in terms of why calendars ought to work better. Well, they ought to provide more retrieval cues in comparison to standardized interviewers. Do they, in fact, do so? And, if they do so, does it matter? That is, if those cues that are made more available in calendars, if they do occur, do they actually matter in terms of data quality?

We’re going to get into the nitty gritty of behavior coding process and the different kinds of behaviors that were observed. So, this is an excerpt example of labor history from that calendar, that life course calendar that I showed you that was computerized. So, we’re somewhere in the middle of this labor history and the interviewers says, “And how long did you stay there, please?” It’s, sort of an interesting, if you will, this is flexible, not standardized.
“How long did you stay there, please?” It’s a bit interesting that it almost sounds like a residence question, but it’s not. Talking about staying at a place of work. And it’s also asking for a duration. How long? Curiously, the respondent provides a timing response, provides a stop time, “October of ’92.” “Okay. And then, around October of ’92, did you take another job?”, a sequential probe. “I took another job and it lasted for, like, a month.” That’s a duration response, a month. “And then, I went to work someplace else.” You’re interested in – “I went to work someplace else”, sequential response, “not that month” and then, we’ll take the next. Now, this is going to be confusing to you. It’s a point of clarification from the interviewer. The interviewer’s following with the respondent was already understanding and that is that spells of work less than three months were not to be recorded.

“So, okay. The next job was at?”, sequential response, “Let’s see.” Explanation of that response in terms of the employer name. Verification of the employer. Verification agreement. “And when did you start working for them?” “In ’92.” You notice that ’92 appears a directive query? Something that’s standardized interviewing would say you shouldn’t do, you shouldn’t direct a respondent to a particular response. But, then it makes sense with regard to this person only having been at a place in a month, that started in October, just really gives you an indication that the interviewer’s really listening to the respondent. “Yes, December of ’92 until May of ’93.” And you can, sort of, get the idea that the respondent’s picking up on their task in terms of now providing the entire spell in terms of its beginning and ending. “Alright. December of ’92 until May of ’93 and then, in May of ’93, did you”, sequential probe, and then you get this fairly rich in terms of interviewing, a history from the respondent. “I went to work and stayed three” – that’s a parallel response, a residence domain, for employer 11 and a data element response. “But, I’m trying to think how long you worked there.” “I came in state 3 in August.” Parallel again. “I have gone – I must have gone for a job”, sequential. “I guess I started in January of ’94”, timing, “and only worked for six weeks duration and then, I went on disability.” Another parallel response.

So, you get a sense that in terms of this calendar format, the nature of the conversation is one in which the language used is in more of a story format. We talked about autobiographical memory as providing stories. It’s more in a story format consisting of all these elements associated with cuing mechanisms.
If we do look at the main question of whether or not it makes a difference between calendar and standardized interviewing, in terms of the prevalence of different kinds of behaviors, indeed, it does. Parallel and sequential retrieval probes by interviewers, retrieval strategies, spontaneous and response are all more prevalent in calendars. Conversational behavioral seeking to clarify meanings are more prevalent, but also, then, we have these more prevalent potentially biasing behaviors, director probing and what we term as unacceptable feedback, which I can explain if anybody wants to know about that later on.

If we look at associations with data quality, it’s not as clear as we would have liked. One of the factors that is important to take into account is people’s experiential difficulty. That is, how complicated are their histories? Did they have multiple jobs in their past? Did they have multiple marriages in their past? And if we look at experiential difficulty, which means a much more difficult retrieval task, the use of these cues leads to greater accuracy, is associated with greater accuracy. It’s probably an overstatement to say it leads to greater accuracy. However, if the past is unremarkable, a higher number of these retrieval probes and strategies is actually associated with less accuracy. Conversational behaviors lead to mixed results and rapport behaviors are also lead to mixed results. And so, we probably have to get even further down into the nitty gritty of understanding, perhaps, in a more qualitative sense what’s going on in these interviews to make some sense out of that.

Paradata. We have recently acquired para data from the American Time Use Survey. There are a number of different variables from these – those auto trails as Para data, which we think will be useful. One of them, as I talked about before, which we actually saw in the calendars, is that people, sometimes, prefer to report in elapsed times in terms of durations and, other times, in timing, in beginning and stop times. That may be important.

Other kinds of things is – well, there are not people that had entered a verbatim or something that was precoded; whether or not the number of activity entries in the paradata is greater than public use data. We have access to public use data as well, which means that there has been some editing going on. Some of those activities were, for some reason, other word deleted and so forth.

We also are able to look at data quality variables, such – and these exist in the public release file or can be extracted from them, such as answer to vague categories; respondent has an unfilled gap in
time; they may provide an overabundance of rounding their answers in terms of timing or in terms of durations and missing key reports of things that you think that people would do every day, such as sleeping, grooming and eating.

We have some preliminary results and, just to, sort of, for the sake of time, move fairly quickly over this, in terms of associating among interview entries, associations between entries and data quality and association between interviewer characteristics and data quality, it’s kind of curious with regard to educated respondents having increased vagueness. You would think that they had always had more decreased vagueness. I, sort of, wonder whether or not there might be more sensitive information that these more highly educated respondents may not want to report on.

[Laughter]

I hadn’t thought about that before this conference, so that was sort of helpful.

I want to illustrate to you now the big picture. And the big picture is moving towards a greater reliance on computerization. This conceptual model was developed by a colleague at the University of Nebraska, Computer Sciences, link it, so I want to give him full credit for this conceptual model. I did some editing to the point that I could understand it. So, for the sake of, perhaps, not being as sophisticated as one would like, in terms of interesting concepts, I’m going to do my best to go through all of this with you in the brief amount of time. I think I only have a couple of minutes. Is that right, Jon?

Well, we do have lots of source of potential information. Para data, behavior coding, interviewer characteristics and respondent characteristics. I talked about behavior coding with calendars, Para data with the American Time Use Survey. Ideally, you could have both. I did mention respondent characteristics to interviewer characteristics as well. And you can then go through various steps in terms of developing more intelligent, smart instruments instead of reliance, if you will, on the dumb instruments that we have now. Data processing, doing such things with regard to, mundane things, perhaps, with regard to the Para data of the ATUS, for confidential reasons, that the verbatims have to be scrubbed out. They have to be sanitized. Well, programs are being developed to do that process instead of relying on the human eye to try to get rid of those verbatims.
Machine learning and pattern recognition. We engage in these kinds of pattern recognition processes and seek to clarify predicting quality, data quality, through our own intuitions. Well, computer scientists are beginning to develop data mining techniques that may be able to do a better job than what our intuitions are able to provide. You have the sense of having a great amount of data with regard to behavior coding. Talk about the richness of transcription data with regard to the Para data itself. Vast amounts of data that, certainly, are going to, pretty much, go beyond our individual abilities to be able to make sense out of them and there might be computer algorithms that actually can do a better job in terms of classifying these various patterns. And this can lead, if you will, to adaptive, assisted instruments that are instruments in which interviewers are still engaged in some way, but which various probes are provided interviewers as to how to deal with the particular respondent. They can be fairly static, such as the need to fill on a gap in a timeline. That’s fairly static. Or they may be much more adaptive in the sense that you may have a history of a different kind of pattern of behaviors that provided indication that a certain error is about to begin. And, hence, you may be able to prevent that error through some sort of intervention through the interviewer.

And, finally, the idea here is to develop an intelligent, self-administered instrument, a virtual interviewer instrument, if you will. That is an instrument that will replace interviewers by being able to examine through the various data mining techniques and the – looking at different classification of behaviors of data and how they predict quality – data quality, to use intelligent agent as an interviewer; in a sense, a virtual interviewer that can produce and tailor various ways of asking respondents questions in a manner in which the best data quality possible and the best memory possible can be derived from that respondent.

And, hence, with regard to a self-administered, web based questionnaire, you really would have a virtual interviewer, if you will, guiding the respondent to maximize the cues that would be available in autobiographical memory. And so, both of those, we would think that we could engage in both of those types of activities in the future with both calendars and time diaries.

So, when I, sort of, just – this is, sort of, all the detail I don’t have time to go through. Why don’t I just, sort of, stop there with regard to the summary and I’m sure that you can, sort of, get a sense of where I’m going with this – importance of using temporal information to assist with remembering; the value of behavior
coding Para data analyses; that this can be used. We can maximize these coming opportunities and self-administrate – and administered questionnaires, especially if you use smart instruments and eventually getting from just assisting interviewers to actually replacing them.

Okay. Thanks.

*Question:* Can you back to the [Inaudible]?

*Robert Belli:* Sure.

*Question:* [Inaudible]?

*Robert Belli:* In the PSID, for example, that life course calendar, it would have been the history of the responses on the PSID. So, that would be the source of validation.

*Question:* Okay. So, they’re remembering answers they gave earlier, to earlier interviewers?

*Robert Belli:* Well, up to 30 years earlier. So, I, you know, they’re not remembering their answering. They’re just, you know, you’re checking the answers they provide in the calendar where the 30 year retrospective report to what they reported up to 30 years earlier in the PSID.

*Question:* Okay.

*Robert Belli:* Al, you had your hand up first.

*Question:* Are you envisioning the smart agents as being avatars or will, that is to say, a smart agent that can ask – literally ask questions using the full range of available, you know, available technology or would this be something – I’m just thinking in terms of would it be restricted to those who are sufficiently literate to be able to fill it out or could it be extended?

*Robert Belli:* Well, that’s an excellent question. I think, of course, at first, you probably would have to have more simple systems; one that would assume literate respondents. But, if we’re looking towards the future and if we’re looking towards, ideally, voice recognition systems, voice recognition systems that can be engaged and automatic and coding of the meaning of the interactions, one could, you know, sort of envision a future in which you do have avatars that provide that kind of virtual human interface and you can have
a very intelligent system in the background that is operating on the basis of its past experiences in terms of what kinds of patterns and sequences lead to better quality replies. And engaging in that kind of interaction with the respondent just using audio exchanges. Yeah.

Question: Yeah, actually, that was my– somewhat, my same question. But, I was thinking that, as we were talking earlier about, you know, recording interviews and people say, “Oh, my God. That’s” and all that. Well, and this is the time to be recording interviews ‘cause I think that the very data that we need to make that kind of decision for artificial intelligence in terms of using voice interviewing and machine based could be based on a database of recorded interviews that are now. And I think that research in that direction should start now because we’re not far off from the technology. And maybe an interim step is to move these kinds of things, obviously, onto tablets and so on as opposed to the little bit more rigid looking web stuff that we see in the American Time Use Survey. But, I think that those are the areas that need funding, but the most basic one, I would say, is to create a database to start to develop, you know, a machine Avatar type of interviewing process that has the appropriate probes, etc., because we could end up with total garbage too. I mean, if you ever tried to use some of these things, you get absurd, you know, answers and that with some things today; but, that’s today and I think, you know, the purpose of this group and the need for the funding is to look further out on the horizon. And I think the time is now to start doing that.

Robert Belli: Yeah. And I fully agree with that idea. I think that we first have to overcome certain impediments and if we’re, sort of, thinking about the future problems, this impediment is really much more minor than that, but it’s more immediate. And that impediment has to do with confidentiality concerns. When you have rich, verbal data, there’s always going to be information there that can easily identify who people are. And, hence, that data has to be scrubbed in some way and developing systems that can scrub data is important. We’ve done that here, if you will, ourselves, using a transcriptioners, okay? Transcriptions and [inaudible]. We’re able to do that. But, if you’re talking about lots and lots and lots of interviews, you can see that relying on human beings to do that becomes very, very labor intensive.

So, developing systems in which we have assurance of being able to get these – this information out in a way that doesn’t compromise confidentiality is a real important concern. It’s even
more of a concern for the federal statistical agencies. I don’t fully understand the reasons for this, but they are unwilling to release transcripts that have actually been de-identified. It will not go through census disclosure because those who have the say as to what information can be released, won’t release this information. The idea here is that the informed consent involves the data being used only for statistical purposes and not, if you will, for these kinds of methodological purposes. That’s my basic understanding. And so, that’s another issue that has to be overcome. So, I agree with you fully, but there are, if you will, some initial constraints that we have to deal with.

Question: Just a follow up to that one. We just talked about it earlier about getting consent for different things. So, maybe a project with that particular focus that had, you know, served dual purposes in which you had a consent dimension involved in it, would allow us to begin to develop such a database ’cause it doesn’t all have to be one survey at one time. It just needs to be an accumulation of data that represents the kind of interviewing that we’re interested in conducting.

Robert Belli: Okay. Thanks.