

GREETING AND RESPONSE: PREDICTING PARTICIPATION FROM THE CALL OPENING

NORA CATE SCHAEFFER*

BO HEE MIN

THOMAS PURNELL

DANA GARBARSKI

JENNIFER DYKEMA

Although researchers have used phone surveys for decades, the lack of an accurate picture of the call opening reduces our ability to train

NORA CATE SCHAEFFER is Sewell Bascom Professor in the Department of Sociology and Faculty Director of the University of Wisconsin Survey Center, University of Wisconsin-Madison. BO HEE MIN is doctoral candidate in the Department of Sociology, University of Wisconsin-Madison. THOMAS PURNELL is Professor in the Department of English, University of Wisconsin-Madison. DANA GARBARSKI is Assistant Professor in the Department of Sociology, Loyola University, Chicago. JENNIFER DYKEMA is Senior Scientist/Survey Methodologist with the University of Wisconsin Survey Center, University of Wisconsin-Madison.

We thank the participants in the Wisconsin Longitudinal Study for their generous contributions of time and information over many years. We would like to thank the reviewers for their extremely creative and thoughtful comments and the editors for many improvements. Mark Banghart and Russell Dimond of the Social Science Computing Cooperative provided essential statistical advice; Ellen Dinsmore assisted with the literature review; Trent Buskirk gave helpful comments on an earlier draft; and Douglas W. Maynard provided helpful guidance to the literature about conversation analysis.

An earlier version of this work was presented at the 2014 annual meeting of the Midwest Association for Public Opinion Research in Chicago and the 2015 annual meeting of the American Association for Public Opinion Research.

This work was supported by grants from the National Science Foundation (SES-1230069) and from the University of Wisconsin - Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation to NCS. Other support for the construction of the data file, analysis, and collection of the data was received from the National Science Foundation (SES-0550705) to Douglas W. Maynard, the Wisconsin Center for Demography and Ecology (National Institute of Child Health and Human Development Center Grant R24 HD047873), the Wisconsin Center for Demography of Health and Aging (National Institute on Aging Center Grant P30 AG017266), the University of Wisconsin Graduate School Research Committee (to Maynard), the William H. Sewell Bascom Professorship, and the University of Wisconsin Survey Center (UWSC). The opinions expressed are those of the authors.

*Address correspondence to Nora Cate Schaeffer, University of Wisconsin Survey Center, University of Wisconsin-Madison, 475 No. Charter St., Madison, WI 53706; E-mail: Schaeffe@ssc.wisc.edu.

interviewers to succeed. Sample members decide about participation quickly. We predict participation using the earliest moments of the call; to do this, we analyze matched pairs of acceptances and declinations from the Wisconsin Longitudinal Study using a case-control design and conditional logistic regression. We focus on components of the first speaking turns: acoustic-prosodic components and interviewer's actions. The sample member's "hello" is external to the causal processes within the call and may carry information about the propensity to respond. As predicted by Pillet-Shore (2012), we find that when the pitch span of the sample member's "hello" is greater the odds of participation are higher, but in contradiction to her prediction, the (less reliably measured) pitch pattern of the greeting does not predict participation. The structure of actions in the interviewer's first turn has a large impact. The large majority of calls in our analysis begin with either an "efficient" or "canonical" turn. In an efficient first turn, the interviewer delays identifying themselves (and thereby suggesting the purpose of the call) until they are sure they are speaking to the sample member, with the resulting efficiency that they introduce themselves only once. In a canonical turn, the interviewer introduces themselves and asks to speak to the sample member, but risks having to introduce themselves twice if the answerer is not the sample member. The odds of participation are substantially and significantly lower for an efficient turn compared to a canonical turn. It appears that how interviewers handle identification in their first turn has consequences for participation; an analysis of actions could facilitate experiments to design first interviewer turns for different target populations, study designs, and calling technologies.

KEYWORDS: Hello; Identification/recognition; Interaction; Nonresponse; Survey introductions.

1. INTRODUCTION

Although survey researchers have used phone surveys for decades, we lack an accurate picture of the opening of the call, and this reduces our ability to train interviewers to succeed from the beginning of the contact. In this study, we use features of the first two turns of the call to predict whether or not a sample

This research uses data from the Wisconsin Longitudinal Study (WLS) of the University of Wisconsin-Madison. Since 1991, the WLS has been supported principally by the National Institute on Aging (AG-9775, AG-21079, AG-033285, and AG-041868), with additional support from the Vilas Estate Trust, the National Science Foundation, the Spencer Foundation, and the Graduate School of the University of Wisconsin-Madison. Since 1992, data have been collected by the University of Wisconsin Survey Center. A public use file of data from the Wisconsin Longitudinal Study is available from the Wisconsin Longitudinal Study, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI 53706, and at <http://www.ssc.wisc.edu/wlsresearch/data/>.

member will participate in a telephone survey. We consider two types of components of each turn: acoustic-prosodic components (such as pitch) and interviewers' actions. We begin with the sample member's first turn, "hello." The prospect of making predictions from the sample member's "hello" is tantalizing: (1) Some contacts with sample members provide little information about the sample member other than "hello," so analysts might like to exploit any information "hello" conveys. (2) The "hello" could potentially provide, for all sample members who answer the phone, information about propensity to participate that has not been influenced by the interviewer, and this information could be used to manage field efforts and measure response propensity in analysis. (3) If the sample member's "hello" provides cues about response propensity, interviewers might be trained to use these cues appropriately.

We then consider the interviewer's initial opportunities for "tailoring." Although "tailoring" originally referred to "changes in interviewer behavior...shaped by real concerns revealed by householders" (Groves and Couper 1996; Couper and Groves 2002), it has been broadened to include other types of responsiveness, including the exchange of greetings (Groves and Benkí 2006; Schaeffer, Garbarski, Freese, and Maynard 2013). We examine the other actions in the interviewer's first turn, which concern "identification/recognition" (Schegloff 1979) and combine self- and institutional identification and a request to speak to the sample member. In the first turn, the interviewer can display competence in projecting and meeting (1) an answerer's plausible concern with the caller's identity and purpose and (2) a plausible expectation that the caller will address these issues (Schegloff 1979) and thereby prevent identification becoming a concern for the answerer and a matter for repair.

We build on earlier investigations but differ in (1) recognizing that actions of the interviewer in the first turn are so structured that the turn as a whole must be considered, (2) documenting the limited structures interviewers actually use in their first turn, (3) comparing turn structures that do ("canonical") and do not ("efficient") accomplish identification, (4) using an analytic sample that includes sample members regardless of where they exit,¹ and (5) predicting participation from features of the turn of each actor that is least affected by the other. We aim for findings with practical implications and to provide grounding for future experiments about how to begin the call by identifying components of opening turns.

We use the Wisconsin Longitudinal Study (WLS), a panel study of those who graduated from high school in Wisconsin in 1957. We examine digital audio recordings from the 2004–2005 wave, when participants were approximately 65 years old. We expect that the greetings and actions of the sample members will reflect the following: expectations for those of their background

1. For example, of our 257 declinations, 89 declined immediately after the turn with the interviewer's identification and a total of 158 declined before the request for participation. Sample members who continue long enough to hear attempts at persuasion are a select group (e.g., Sturgis and Campanelli 1998; De Leeuw and Hox 1996).

and cohort (e.g., about how a stranger who is calling should address them); experience with prior rounds of the WLS (most recently 1992–1993 for most); review of the advance letter in the current wave (for most); and the sample member’s observation of attempts to contact them on caller ID or answering machine messages (for some). It is consequential for the interaction that the interviewer can ask for the sample member by name and does not need to select someone from the household.

Our sample, design, and analytic approach could limit or strengthen generalizations. If the content or structure of the turns we study occur only with this study design or population, then our results might be most relevant for panel studies in which sample members can be asked for by name or for studies of older adults—of which there are important instances.

2. BACKGROUND AND MOTIVATION

Because the motivation of hypotheses is somewhat different for the sample member’s “hello,” the interviewer’s greeting, and the actions in the interviewer’s first turn, we introduce each separately.

2.1 The Sample Member’s “Hello”

We ask whether the sample member’s “hello” forecasts the outcome of the call. “Hello” is highly conventional (Schegloff 1986) but may communicate nonetheless. For example, if the sample member does not know the caller’s identity or reason for calling, their “hello” may communicate that. There is evidence that speakers project stances and relationships with listeners (e.g., Schegloff 1998; Pillet-Shore 2012; Kockelman 2004) and that listeners perceive these and other characteristics.² Drawing on Pillet-Shore’s (2012, p. 383) analysis of how greetings display stance in face-to-face interactions, we hypothesize that the following features of a “large” greeting will predict participation: longer duration, higher pitch (the best operationalization we have available for “smile voice”), a pattern of falling pitch (pitch pattern), and wider pitch span.

2.2 First Opportunity for “Tailoring”: the Interviewer’s Greeting

Unlike our hypotheses for the sample member’s greeting, which focus on its absolute qualities, our hypotheses about the interviewer’s greeting focus on its

2. Listeners make varied (reliable or accurate) judgments based on small acoustic samples (e.g., Banse and Scherer 1996; Dykema, Diloreto, Price, White, and Schaeffer 2012; McAleer, Todorov, and Belin 2014; McCulloch 2012; McCulloch, Kreuter, and Calvano 2010; Purnell, Idsardi, and Baugh 1999; Scharinger, Monahan, and Idsard 2011; Scherer, Banse, Wallbott, and Goldbeck 1991; Schweinberger, Kawarhara, Simpson, Skuk, and Zaske 2014; Tartter and Braun 1994).

responsiveness, although we report findings about both. We hypothesize that a responsive greeting by the interviewer will increase the likelihood of participation, for example, by displaying competence as an interactional partner. In acoustic terms, a responsive greeting could either mirror or complement. The literature does not provide guidance about the forms of acoustic tailoring, so we explore several. The interviewer's first turn also offers an opportunity for lexical tailoring: With the WLS cohort, we expect the reciprocal "hello" to be more successful than the standard casual greeting, "hi," used by many interviewers.³

2.3 Actions in the Interviewer's First Turn

The interviewer's first turn begins with a greeting and continues until the sample member speaks again. As described in our interactional model of the recruitment call (Schaeffer et al. 2013), the interviewer's first turn potentially includes a number of crucial actions. A "canonical" first turn for the interviewer would look much like the sample script that appeared on the screen. The script included greeting, self-identification, institutional identification, and request to speak to the sample member; interviewers were trained to use first and last names: *Hello. My name is (NAME). I am calling from the University of Wisconsin Survey Center at the University of Wisconsin-Madison. May I please speak to (NAME)?* Interviewers were authorized to adapt the script to sound more conversational (Morton-Williams 1993; Houtkoop-Steenstra and van den Berg 2002). When a sample member was called to the phone by a third party who answered the call, a canonical turn included a greeting, self- and institutional identification by the interviewer, and an optional acknowledgement or confirmation by the interviewer of the sample member's identity.

We use several perspectives to predict consequences of the construction of the interviewer's first turn. First, a call recipient may expect a stranger who is calling to identify themselves in their first turn (Schegloff 1979). Such conventions help manage social exchange, identification, footing, and such. The predictability of conventional practices lets participants assess each other's interactional competence and, perhaps, make other inferences. Second, social exchange theory suggests that by offering identity in their first turn the interviewer (1) generates an obligation for the sample member to confirm their identity in return and (2) builds trust (Gouldner 1960; Dillman 1978; Dillman, Smyth, and Christian 2014). Finally, "footing" (Goffman 1979) describes how speakers and listeners align; the everyday concept of "footing" refers to the basis of information or trust on which an interaction proceeds. The footing of these actors differs: in a list sample or panel study, the interviewer knows the name, telephone number, and other facts about the sample member, but the sample member has no information about the interviewer.

3. Schaeffer et al. (2013) report this comparison with a slightly different operationalization.

In a canonical introduction, the interviewer completes “identification/recognition” and then asks for the sample member; this makes the sample member’s confirmation of their identity an act of reciprocity. By contrast, in an “efficient” introduction, the interviewer first verifies that they have reached the sample member. This “efficiency” conspicuously betrays the interviewer’s privileged knowledge, establishes an unequal footing, and may make the interviewer’s interactional competence questionable. Thus, we expect lower likelihood of participation if the interviewer begins with an efficient turn. This implies that we do not expect individual actions—such as asking to speak to the sample member—to have the same effect regardless of how the turn is constructed.

We focus on actions, but we are able to examine other qualities of the interviewer’s first turn. Opportunities for politeness in the first turn are limited, but we expect polite turns to be more successful, particularly with the WLS cohort. A polite turn acknowledges (1) the sample member’s power in the interaction by mitigating the interviewer’s request (e.g., “please” and mitigating language like “may I”) and (2) the social distance between the actors (e.g., use of titles and polite words) (Brown and Levison 1987; Holtgraves and Yang 1992; Stephan, Liberman, and Trope 2010). (The conventions for acknowledging relative power and mitigating a request probably vary for different populations.) To complete our analysis of the first turn, we include measures of disfluency (e.g., Conrad, Broome, Benkí, Kreuter, Groves, et al. 2013) that may affect a sample member’s perception of the interviewer as a competent interactional partner.

2.4 Previous Research

Most previous research about acoustic or perceived properties of speakers during the opening of the recruitment call has focused on the interviewer and not specifically on “hello” (e.g., Oksenberg and Cannell 1988; Oksenberg, Coleman, and Cannell 1986; van der Vaart, Ongena, Hoogendoom, and Dijkstra 2006; Groves, O’Hare, Gould-Smith, Benkí, and Maher 2008; Conrad et al. 2013). For example, Benkí, Broome, Conrad, Kreuter, and Groves (2011) considered the interviewer’s average median pitch and variability in pitch over the first 13 turns, not just “hello.”

Two analyses examined “hello” with a study design quite different from ours. Groves and Benkí (2006) found that the relationship between the rated “friendliness” of the householder’s “hello” and the likelihood of an interview, appointment, or callback was in the predicted direction but was not significant. For the interviewer’s first turn, they examined acoustic properties, but not actions. In later work, Benkí, Broome, Conrad, Groves, and Kreuter (2013, p. 13) compared “pitch change” for “hello” (using an operationalization that incorporated information after the first turns) for answerers and interviewers within different outcome groups. Our studies differ in operationalizations (we use only information in the first turn of each actor) and analytic approach (we predict outcome from the first turns), so our results are difficult to compare.

With respect to the impact of the interviewer's actions, [Campanelli, Sturgis, and Purdon \(1997\)](#) reported that participation is more likely when interviewers introduce themselves in face-to-face interviews, but they do not examine where the "introduction" is located or the structure of the first turn. [Maynard, Freese, and Schaeffer \(2010\)](#), [Schaeffer et al. \(2013\)](#), [Maynard and Hollander \(2014\)](#), and [Nolen and Maynard \(2013\)](#) analyzed various actions and features of action during the recruitment call for WLS but did not focus on the first turns.

In summary, we examine whether acceptance is associated with (1) a "large" greeting or other acoustic properties of the sample member's "hello" or (2) the acoustic properties and possible acoustic or lexical reciprocity of the interviewer's greeting. We then consider whether acceptance is less likely when the interviewer uses an efficient first turn in which they do not identify themselves; we also look at other features of the turn, such as its politeness.

3. DATA

3.1 Sample

We use digital recordings from the 2004 round of the Wisconsin Longitudinal Study. WLS began with a one-third sample of 1957 Wisconsin high school graduates who were followed in the intervening decades: 1964 (mail to parents), 1975 (telephone), 1992 (telephone and mail), and 2004 (telephone and mail). Responses to the main mode of data collection during follow-up were 87, 90, 87, and 80 percent of those who were still living, respectively. When original sample members known to be deceased are included in the denominator, the 2004 round interviewed 70 percent of the original sample. We have considerable information about all sample members fielded in 2004 and audio recordings of contacts with the sample member by the interviewer.

We use information from the WLS ([Hauser 2005](#)) to construct a case-control study. We constructed 257 pairs of cases (the maximum number of pairs we were able to make). In the first contact with a WLS interviewer, one pair member declined to be interviewed and the other pair member accepted. Pair members are matched on gender, past participation, and estimated propensity to participate.⁴ For the analysis of actions, we use all 257 pairs. For the

4. The impact of clustering within interviewer is limited by the large number of interviewers in our analytic sample compared to the number of sample members. We have 138 interviewers, and the mean number of cases per interviewer is about 3.7 for both acceptances and declinations. Analytically, we expect that interviewer effects would be conveyed primarily via the interviewer's actions, actions that are usually unobserved but that we are able to measure. [Schaeffer et al. \(2013\)](#) give details about the sample, estimated propensity scores, matching, and reliability of coding of actions. The model estimating the propensity to participate included education, high school class rank, high school cognitive assessments, self-reported health, sex, and past participation. In addition to being matched on estimated propensity to participate, pairs were matched on gender and past participation to try to control influences on current participation. Details about response

acoustic analysis, we drop a pair if one sample member in the pair did not say “hello” or one sample member’s greeting token was too poorly recorded to analyze. Of the 514 cases, 436 have usable “hello” recordings from the sample member; after eliminating pairs in which one sample member did not have a usable recording, 187 pairs (374 cases) remain. Because of the case-control design, the analytic sample is not a probability sample of the larger WLS sample, and calculations from our analytic sample (e.g., frequencies of a particular action) do not describe the WLS sample more generally.

We are interested in the consequences of each actor’s first turn. In most calls, the sample member answers the telephone. A third party answers the telephone and calls the sample member to the telephone in 95 of the 374 calls in the acoustic analysis and 135 of the calls in the full analytic sample of 514 cases. For these “third-party calls,” we use the sample member’s greeting when they come to the telephone and the interviewer’s subsequent first turn. We discuss later how these calls differ from those in which the sample member answers.

3.2 Greeting Tokens and Acoustic Measures

The acoustic analysis includes only pairs in which the sample member began with “hello” (over 94 percent of the sample). Interviewers’ greetings were more variable, and many used “hi.” Measures analyzed include pitch (mean, minimum, and maximum pitch [Hz]); pitch span (Hz); pitch pattern; duration of each actor’s greeting; and the latency between the end of the sample member’s greeting and the beginning of the interviewer’s turn (see [table 1](#)). Our project is necessarily exploratory, and many of our measures of pitch or duration are correlated. Because we lack *a priori* justification for specific measures of acoustic reciprocity, we examine several (correlated) possibilities: mirroring (e.g., both in the upper, both in lower, or both in the same extreme of their respective distributions) or complementarity (e.g., one in each extreme). This lets us assess whether our findings depend on details of the operationalizations and identify the most interpretable version. We examine lexical reciprocity by comparing “hello” to other greeting tokens by the interviewer.

3.3 Standardization and Adjustment

Our method of standardizing measures of pitch and duration adopts the point of view of the participants. We speculate that interviewers would compare the sample member’s “hello” to that of other adults of the same age and gender, and we use the sample members to approximate this comparison group. We apply the same logic for the comparisons made by the sample members

rate can be found at (http://www.ssc.wisc.edu/wlsresearch/documentation/retention/cor1004_retention.pdf). All interviews were conducted in English, most on a landline.

Table 1. Summary of Acoustic Measures^a

| Property | Actor ^b | Concept | Measurement | Notes about analytic variable |
|----------|--------------------|---|--|---|
| Pitch | SM & INT | Pitch of greeting token (mean, minimum, or maximum) | Mean, minimum, or maximum fundamental frequency of the greeting token ("hello" for SM, "hello" or "hi" for INT) in Hertz. | Each measure standardized using mean and standard deviation of other sample members of same gender. |
| | SM & INT | Pitch span of greeting token | Maximum and minimum fundamental frequency of the greeting in Hertz. | Computed as maximum frequency of the greeting token divided by the minimum frequency. Span of greeting token was the minimum-maximum ratio converted from Hertz to semitones. |
| | SM & INT | Pitch pattern of greeting token | The pattern of rising, falling, or constant pitch during the delivery of the greeting token. | Comparison across these categories (e.g., falling versus all others). |
| Duration | SM | Duration of greeting token | Duration of the greeting token in seconds. Boundaries of the token ("hello") were identified. Duration is the time between the boundaries. | Standardized using mean and standard deviation of other sample members of same gender. Duration of entire token was used (rather than just the final vowel, /o/) to allow for analysis that included interviewers who say "hi." |
| | | | | INT |

(continued)

Table 1. continued

| Property | Actor ^b | Concept | Measurement | Notes about analytic variable |
|----------|-----------------------------|----------------------------|--|---|
| | | Duration of greeting token | Duration of the greeting token in seconds. Boundaries of the token ("hello" or "hi") were identified. Duration is the time between the boundaries. | Because "hi" and "hello" are of different lengths, the duration was first adjusted by the ratio of the mean duration of "hello" to the mean duration of "hi" for interviewers of the same gender. The adjusted duration was then standardized using the mean and standard deviation of other interviewers of same gender. |
| INT | Latency as transition delay | | Time in seconds between the end of the sample member's last utterance in the response-to-summons turn and the onset of the interviewer's subsequent turn. Latency ends with first utterance from the interviewer, even if that utterance is a token. Measured in Audacity. | Standardized using mean and standard deviation of other sample members of same gender. |

^aTechnical details for all variables are in the online Supplementary materials. Acoustic variables measured in Praat (Boersma and Weenink, 2012, <http://www.fon.hum.uva.nl/praat/>).

^b"SM" indicates "sample member"; "INT" indicates "interviewer."

(although without as strong a justification). For duration we also standardize within actor and gender, and for interviewers we first adjust to make “hello” and “hi” comparable. (Details about adjustments and standardization are in [table 1](#) and the [Supplementary materials online](#).) These procedures let us examine the qualities of the greeting regardless of the type of greeting or actor. We operationalized reciprocity similarly for both pitch and duration by examining the relative positions of the actors in the distribution, for example, both in the top third of that actor’s distribution of pitch.

3.4 Interviewer’s Actions

The coding of actions in the interviewer’s first turn extended codes previously developed ([Schaeffer et al. 2013](#); [Maynard and Hollander 2014](#)). [Table 2](#) summarizes these measures, some of which are complementary or dependent in other ways.

3.5. Analysis

The analysis uses bivariate conditional logistic regressions of participation on the individual independent variables. For each dummy variable, the comparison is to all other cases in the analysis. As a result, some contrasts are not independent of each other, but our approach is exploratory and allows for flexible description of the results. We used a conditional logit (clogit in Stata). The following likelihood function for clogit with groups (that is, pairs of observations) is based on [Chamberlain \(1980\)](#)⁵:

$$L = \sum_{\{i \in I_1\}} \left(\sum_{\{j: y_{ij}=1\}} \left[(\mathbf{x}_{i2} - \mathbf{x}_{i1}) [(-1)^{I(j=2)} \boldsymbol{\beta}] - \ln \left(1 + e^{(\mathbf{x}_{i2} - \mathbf{x}_{i1}) [(-1)^{I(j=2)} \boldsymbol{\beta}]} \right) \right] \right),$$

where

- i is the group identifier;
- ij , where $j \in \{1, 2\}$, is the j th observation of the i th group;
- $I_1 = \{i | y_{i1} + y_{i2} = 1\}$;
- x_{ij} is the row of covariates associated with the j th observation of the i th group;
- $I(j = 2)$ is the indicator function for $j = 2$.

5. The likelihood function minimized by clogit is described on the Stata clogit page (<http://www.stata.com/manuals14/rclogit.pdf>). This section refers to several other sources, including [Chamberlain \(1980\)](#), which is the basis for the likelihood function above (Mark Banghart, personal communication). The first beta is a multiplier to the difference in the x values in the i th group. The bold font for the x and betas in the formula represents that there may be more than one regressor in the model.

Table 2. Concepts and Operationalizations for Actions in Interviewer's First Turn^a

| Panel A. Construction of interviewer's turn | |
|---|--|
| Concept | Conceptual definition |
| Efficient turn: strict | <p>This structure confirms sample member's identity efficiently but displays unequal information footing of actors and delays identification/recognition. Interviewer asks to speak to sample member without self-identifying.</p> <p>A third party^c answers and calls the sample member to the phone.</p> |
| Efficient turn: variants | <p>This structure confirms sample member's identity efficiently but displays unequal information footing of actors and delays identification/recognition. Interviewer confirms sample member's identity. In the few cases in which turn includes one form of identification, it also includes an intrusive action^d that displays unequal footing of actors.</p> <p>The sample member answers.</p> |

Actions in interviewer's first turn after sample member greeting^b

Greeting + request to speak to sample member
Example: "Hello. May I please speak to Mr. Smith?"

Greeting + at least one of these actions: address to sample member in greeting, confirmation of sample member's identity
Example: "Hello. Is this Mr. Smith?" or "Hello, Mr. Smith."

Greeting + at least one of these: self-identification, institutional identification + confirmation of sample member's identity
Example: "Hello. I'm calling from the University of Wisconsin Survey Center. Is this Mr. Smith?"

Type of call

The sample member answers.

The sample member answers.

Actions in interviewer's first turn after sample member greeting^b

| | | |
|--------------------------------|--|---|
| Canonical first turn: strict | <p>A third party^c answers and calls the sample member to the phone.</p> <p>The sample member answers.</p> | <p>Confirmation of sample member's identity</p> <p>Example: "Is this Mr. Smith?"</p> |
| Canonical first turn: variants | <p>This structure performs identification/recognition in the first turn and equalizes information footing between interviewer and sample member. Turn has self- and institutional identifications and request to speak to sample member.</p> | <p>Greeting + self-identification + institutional identification + request to speak to sample member</p> <p>Example: "Hello. My name is Emily Jones. I'm calling from the University of Wisconsin Survey Center. May I please speak with Mr. Smith?"</p> |
| Canonical first turn: variants | <p>A third party^c answers and calls the sample member to the phone.</p> <p>The sample member answers.</p> | <p>Greeting + self-identification + institutional identification + <i>one of these actions: address to sample member in greeting, confirmation of sample member's identity</i>^b</p> <p>Example: "Hello. This is Emily Jones calling from the University of Wisconsin Survey Center. Is this Mr. Smith?"</p> <p>Greeting + any two of these actions: self-identification, institutional identification, request to speak to sample member</p> <p>Example: "Hello. I'm calling from the University of Wisconsin Survey Center. May I please speak with Mr. Smith?"</p> |

(continued)

| Concept | Definition | Measures | Actions and qualities of actions counted |
|------------|---|--|--|
| Politeness | Polite elements acknowledge the social distance between actors and the sample member's power in the interaction and mitigate the request. | Polite first turn: number of polite elements in first turn | <p>Greeting is polite: "Hello" OR "Good morning/afternoon/evening"</p> <p>Request to speak to sample member is mitigated by asking permission: "May I speak to?"</p> <p>Request to speak to sample member includes "please"</p> <p>Self-identification uses "My name is" rather than "This is" Self-identification uses full name:</p> <p>"My name is <first and last name>" OR "This is <first and last name>"</p> <p>Address to sample member uses last name in greeting and request to speak to sample member</p> <p>Address to sample member uses title: "Ma'am/Sir" OR "Mr./Mrs./Ms." in greeting and request to speak to sample member</p> |

| | | |
|------------|---|---|
| | Polite greeting: number of polite elements in greeting | See “number of polite elements in first turn” |
| | Very polite first turn: interviewer incorporates a polite element in 4 or more locations (out of 6 possible locations in up to 3 actions) | See “number of polite elements in first turn” |
| Disfluency | Disfluent speech is characterized by tokens and may communicate that the interviewer is not a competent interactional partner. | Disfluent opening: interviewer’s first utterance is a token or broken-off greeting token |
| | | Disfluent first turn: interviewer’s first turn includes at least one token regardless of location |
| | | See “disfluent opening” |

^aIn a small number of cases (30 out of 514), the interviewer’s first turn took place over more than one turn. In almost all of these 30 cases, the sample member asked for a repetition due to a hearing problem, and the interviewer then restarted the first turn. In a few cases, the sample member issued a token or similar minor utterance and the interviewer continued their turn. In all these cases, the interviewer’s completed turn was evaluated in classifying the case.

^bActions shown in italics sometimes occurred, but their presence or absence did not affect the classification of the interviewer’s turn.
^cFor third-party calls, we considered the interactional context in analyzing the turn construction. Because a third party brought the sample member to the phone, actions in the turn included an acknowledgement of the sample member in the greeting (“Mr. Smith?”) or, in some cases, a repetition of the request to speak to the sample member.

^dThe most common intrusive action was the “sample member identity confirmation” in sample member calls and the “sample member verification,” which required verifying the high school of the sample member, in calls in which a third-party answered. Both actions revealed the interviewer’s privileged knowledge about the sample member. These actions were rare in the first turn, but when present disqualified the turn from being “canonical.”

The outer summation is over all pairs in which the pair's responses contain one 0 and one 1. The inner summation is over the single observation within the pair in which the response is 1.

Conditional logit is similar to a fixed effect logit in which the matching characteristics are used as categorical regressors in the model. The analysis thus adjusts for characteristics that the pairs are matched on and anything else that they have in common. A conditional logistic regression estimates the association between the within-pair action of interest and participation; it “conditions” the intercept for each pair out of the analysis. The intercepts for the pairs are nuisance parameters and not of substantive interest but can bias estimates if not accounted for. Because our sample size is small and we want to identify avenues for future investigation, we report specific p values; we discuss relationships that are significant with the relatively generous $\alpha = 0.10$, but note when results are marginal by conventional standards ($\alpha = 0.05$).

4. RESULTS

For mean and minimum pitch, there are no statistically significant associations between continuous measures for either actor or for indicators of reciprocity by the interviewer and subsequent participation (not shown, every $p > 0.17$), and we do not discuss these measures further. The key prediction for pitch pattern, that falling pitch would predict participation compared to other patterns, is not supported for either actor, nor were our measures of ways the interviewer might reciprocate pitch pattern (i.e., both the same pattern or both opposite; results for pitch pattern not shown, each $p > 0.24$); however, we note that for sample members pitch pattern is less reliable than our other pitch measures (see the [Supplementary materials](#)).

4.1 The Sample Member's “Hello”

[Table 3](#) presents results for the sample member's “hello.” The continuous measure of maximum pitch does not predict participation ($p = 0.21$); but, as predicted, sample members in the upper 30 percent of the distribution (our approximation to “smile voice”) are more likely to participate than those in the lower 70 percent ($OR = 1.69, p = 0.03$). Maximum pitch is also a component of pitch span, but the pattern of results is clearer for the sample member's pitch span: The odds of participation are higher when the sample member's pitch span is greater ($OR = 1.24, p = 0.05$). The results for sections of the distribution are consistent with a linear relationship: those with a pitch span in the upper 30 percent of the distribution have a higher odds of participation than those in the lowest 70 percent ($OR = 1.74, p = 0.02$), and those whose pitch span is in the lowest 30 percent of the distribution have a lower odds of participation

Table 3. Bivariate Conditional Logistic Regressions of Acceptance of the Request to Participate on Features (Pitch, Duration) of the Sample Member’s “Hello”

| Measure | Definition | No. ^b | Odds ratio | <i>p</i> (2-tailed) | 95% CI | |
|--------------------------|--|------------------|------------|---------------------|--------|-------|
| | | | | | Lower | Upper |
| Pitch^a | | | | | | |
| Maximum | Maximum of standardized pitch (continuous) | 374 | 1.14 | 0.21 | 0.93 | 1.41 |
| | Top 30% of maximum pitch (= 1, 0 = all others) | 374 | 1.69 | 0.03 | 1.04 | 2.75 |
| | Lowest 30% of maximum pitch (= 1, 0 = all others) | 374 | 1.24 | 0.33 | 0.81 | 1.90 |
| Span | Span of standardized pitch (continuous) | 374 | 1.24 | 0.05 | 1.00 | 1.55 |
| | Top 30% of pitch span (= 1, 0 = all others) | 374 | 1.74 | 0.02 | 1.08 | 2.79 |
| | Lowest 30% of pitch span (= 1, 0 = all others) | 374 | 0.62 | 0.04 | 0.39 | 0.98 |
| Duration | Standardized duration of greeting token in seconds (continuous) ^a | 374 | 1.06 | 0.57 | 0.87 | 1.30 |

^aMeasures of pitch are standardized using the mean and standard deviation of sample members of the same gender in the sample. See the [Supplementary materials online](#) for details.

^bSample ($n = 374$) includes pairs in which both sample members in the pair said “hello” and had recordings for which acoustic analysis could be conducted.

than those in the upper 70 percent ($OR = 0.62$, $p = 0.04$). The duration of the sample member’s greeting is not associated with participation ($p = 0.57$).

4.2 The Interviewer’s Greeting

Table 4 presents results for the interviewer’s greeting. The continuous measure of maximum pitch is not associated with participation ($p = 0.22$), but interviewers whose pitch is in the top 30 percent of their distribution may have lower odds of participation than those in the lower 70 percent ($OR = 0.64$, $p = 0.07$), suggesting that a greeting with “smile voice” may not be appropriate for a stranger who is calling. There is no evidence that the odds of participation are greater if the interviewer reciprocates the sample member’s maximum pitch by being, or in the same or opposite extreme of the distribution as the sample member (these results not shown, every $p > 0.57$). None of the measures of the interviewer’s pitch span or the way in which it reciprocates the sample member’s pitch span are significant predictors of participation (these results are not shown; all $p > 0.30$).

Table 4. Bivariate Conditional Logistic Regressions of Acceptance of the Request to Participate on Features (Pitch, Duration of Token, Response Latency) of the Interviewer's Greeting Token

| Measure | Definition | 95% CI | | | | |
|----------------------------|---|------------------|------------|---------------------|-------|-------|
| | | No. | Odds ratio | <i>p</i> (2-tailed) | Lower | Upper |
| Maximum pitch ^a | Maximum of standardized pitch (continuous) | 374 ^c | 0.88 | 0.22 | 0.72 | 1.08 |
| | Top 30% of maximum pitch (= 1, 0 = all others) | 374 ^c | 0.64 | 0.07 | 0.40 | 1.03 |
| | Lowest 30% of maximum pitch (= 1, 0 = all others) | 374 ^c | 0.93 | 0.74 | 0.60 | 1.44 |
| | Standardized duration of greeting token (adjusted) in seconds (continuous) | 340 ^d | 1.04 | 0.75 | 0.84 | 1.28 |
| Duration ^b | Both in top 30% of duration of greeting token (= 1, 0 = all others) | 340 ^d | 0.83 | 0.60 | 0.42 | 1.65 |
| | Both in top or both in bottom 30% of duration of greeting token (= 1, 0 = all others) | 340 ^d | 0.63 | 0.09 | 0.37 | 1.07 |
| Latency | Both in bottom 30% of duration of greeting token (= 1, 0 = all others) | 340 ^d | 0.44 | 0.06 | 0.19 | 1.02 |
| | Complementary extremes (versus not) | 340 ^d | 1.00 | 1.00 | 0.57 | 1.76 |
| | Standardized response latency in seconds | 514 ^e | 1.14 | 0.15 | 0.95 | 1.36 |
| | Long latency (1 = longest 30%, 0 = all others) | 514 ^e | 1.41 | 0.08 | 0.96 | 2.07 |
| | Short latency (1 = short 30%, 0 = all others) | 514 ^e | 0.74 | 0.12 | 0.50 | 1.08 |

^aMeasures of pitch are standardized using the mean and standard deviation of interviewers of the same gender in the sample. See the [Supplementary materials](#) for details.

^bDuration is standardized using the mean and standard deviation of the interviewers of the same gender in the sample. In addition, interviewer greetings are first adjusted to account for the different lengths of “hello” and “hi.” See the [Supplementary materials](#) for details.

^cSample includes pairs in which both sample members in the pair said “hello” and had recordings for which acoustic analysis could be conducted.

^dAnalysis omits from sample in footnote “c” pairs in which the interviewer used a greeting other than “hello” or “hi.”

^eAnalysis includes all available analytic pairs because acoustic details for the sample member were not required and no restrictions on greeting were required.

The continuous measure of duration of the interviewer's greeting is not associated with participation ($p = 0.75$). For reciprocity, when the interviewer mirrors either a long or short greeting token from the sample member (versus others), the relationship is marginally significant but not in the predicted direction ($OR = 0.63, p = 0.09$). This finding appears to be driven by the negative effect of reciprocity when both actors provide short greetings ($OR = 0.44, p = 0.06$). It is plausible that a short token from the sample member projects "hurry," but a reciprocation by the interviewer conveys "curt" or "unfriendly."

The continuous measure of the latency between the end of the sample member's greeting and the beginning of the interviewer's is not associated with participation ($p = 0.15$), although interviewers with the longest latency have higher odds of success ($OR = 1.41, p = 0.08$), possibly because they use this time for processing or for "planning" their first turn.

4.3 Interviewers' Actions

Although interviewers were authorized to use a "flexible" introduction, the vast majority of both acceptances (81 percent) and declinations (84 percent) used a canonical or efficient first turn; 95 percent used one of these constructions or the variants. This strong patterning means that we do not have sufficient variation to estimate the impact of each action (e.g., presence or absence of a self-identification) on the outcome.

[Table 5](#) presents results for the interviewers' actions. What the interviewer can accomplish in the first turn depends in part on the cooperation of the sample member; nevertheless, the number of actions in the first turn is not associated with participation ($p = 0.26$). The analysis of turn construction addresses our principal hypothesis. When the interviewer's turn is efficient (compared to canonical and other), the odds of participation are substantially and significantly lower ($OR = 0.65, p = 0.02$ for strict; $OR = 0.69, p = 0.05$ including minor variants). Panel A of [figure 1](#) illustrates how an efficient introduction could affect studies under different assumptions about the base response rate for the study; for example, if a study to which our odds ratio applied would obtain a 50 percent response rate with an equal number of efficient and canonical introductions, the predicted difference in the response rate with an efficient as compared to a canonical introduction would be between 10 and 11 percent.⁶ In our

6. See [Long \(1997, pp. 75–79\)](#). Because our independent variable is categorical, we estimate the change in predicted response rate varying the response rate of the study for which the prediction is being made. Our matched pairs design does not allow us to estimate the relative proportion of, say, efficient and canonical introductions in our sample, so we calculate the estimated difference in their impact on the response rate assuming that we have equal numbers of both. This approach simulates the impact one might see in an experiment in which an equal number of cases were assigned to each type of introduction. We particularly thank the reviewer who suggested the method and citation and Mark Banghart and Russell Dimond, who helped us implement the reviewer's suggestion.

Table 5. Bivariate Conditional Logistic Regressions of Acceptance of the Request to Participate on Actions of the Interviewer in the First Turn

| Measure and definition | No. ^a | Odds ratio | <i>p</i> (2-tailed) | 95% CI | |
|---|------------------|------------|---------------------|--------|-------|
| | | | | Lower | Upper |
| Turn construction | | | | | |
| Number of actions in first turn (1–5) | 514 | 1.11 | 0.26 | 0.93 | 1.32 |
| Efficient turn (= 1, 0 = efficient variants + canonical + canonical variants + other) | 514 | 0.65 | 0.02 | 0.46 | 0.93 |
| Efficient turn and variants (= 1, 0 = canonical + canonical variants + other) | 514 | 0.69 | 0.05 | 0.48 | 0.99 |
| Politeness | | | | | |
| Number of polite elements in first turn (0–9) | 514 | 1.04 | 0.51 | 0.93 | 1.16 |
| Number of polite elements in greeting (0–3) | 514 | 1.23 | 0.20 | 0.90 | 1.70 |
| Greeting includes polite element (= 1, 0 = absent) | 514 | 1.28 | 0.21 | 0.87 | 1.87 |
| Very polite first turn (1 = 5 or more out of 9, 0 = all others) | 514 | 1.75 | 0.07 | 0.95 | 3.23 |
| Greeting token (1 = hello or good morning/afternoon/evening, 0 = all others) | 502 | 1.36 | 0.12 | 0.92 | 2.01 |
| Greeting token (1 = hello, 0 = hi) | 458 | 1.49 | 0.06 | 0.98 | 2.26 |
| Disfluency | | | | | |
| Turn begins with disfluency token (= 1, 0 = absent) | 514 | 0.55 | 0.09 | 0.27 | 1.10 |
| Disfluency token present in first turn (= 1, 0 = none) | 514 | 1.09 | 0.39 | 0.89 | 1.34 |

^a Analysis includes pairs in which both sample members and interviewers had relevant actions.

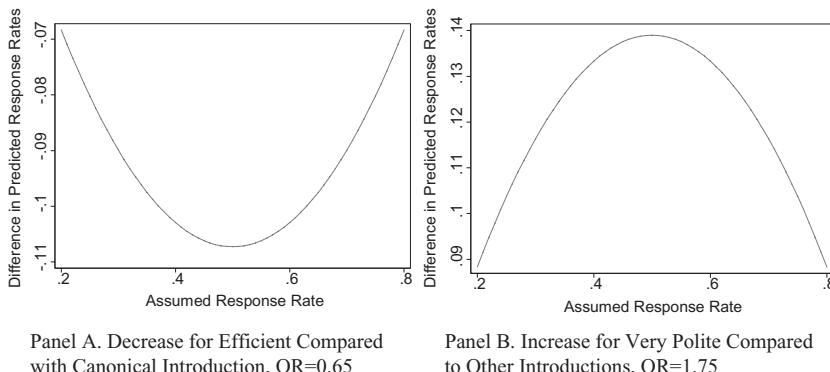


Figure 1. Difference in Predicted Response Rate for Characteristics of Introduction for Values of Response Rate between .2 and .8, Assuming That the Characteristics Are Used with Equal Frequency.

study, if sample members expect identification in the interviewer's first turn, the efficient introduction should lead them to initiate repair with questions such as "Who is this?" or "What is this about?" And when the sample member asks "wh- " questions (in contrast to length-of-interview questions) before the request to participate, the odds of acceptance decrease substantially (Schaeffer et al. 2013).⁷

We examined several operationalizations of politeness; only for the indicator of a very polite first turn are the odds of participation significantly higher (OR = 1.75, $p = 0.07$) (see also Schaeffer et al. 2013). Panel B of figure 1 illustrates the impact of being very polite; if a study to which our odds ratio applied would obtain a 50 percent response rate with an equal number of a very polite and not very polite first turns, the predicted difference in the response rate with a very polite introduction is just under 14 percent. In addition, "hello" is associated with increased odds of participation compared to "hi" (OR = 1.49, $p = 0.06$), perhaps because "hello" reciprocates the sample member's token because "hi" is casual in a way that these older sample members do not like or because "hello" indexes other features of the turn, such as its politeness (see also Schaeffer et al. 2013).

We also examined the implications of disfluency in the interviewer's first turn. Only 25 percent of the first turns in our analytic sample included a disfluency token, and in only 7 percent of the turns was that disfluency in an initial position. The odds of participation are lower if the interviewer begins with a

7. Canonical and efficient calls have different trajectories; nevertheless, the proportion of our cases that exit by key turning points (e.g., before the request to participate) is the same for both. In our analytic sample, "wh- " questions immediately follow the interviewer's first turn in 1.9 percent of cases with canonical (or variant) openings and 6.7 percent of cases with openings that are efficient (or variants; $p = 0.01$, one-sided). "Wh- " questions also occur later, of course.

disfluency token (OR = 0.55 at the marginally significant level of $p = 0.09$),⁸ but are not affected if there is a disfluency anywhere in the first turn ($p = 0.39$).

5. DISCUSSION

Although telephone surveys have been conducted for decades (e.g., [Tourangeau 2004](#)), studies of interaction during recruitment have focused on refusals and the response to them (e.g., [Maynard and Schaeffer 1997](#)). The specific actions in the opening turns, their features, and sequential placement have not been previously described to our knowledge, but interviewers must be trained for this key moment when sample members are contacted by phone.

Our analysis of the sample member’s “hello” emphasizes the positions of the participants in the first moments of the call. Although we could not fully operationalize Pillet-Shore’s “large” greeting (2012), the sample member’s pitch span and a related measure — a relatively high maximum pitch (smile voice) — predicted participation in a way consistent with her analysis; pitch pattern (which was challenging to operationalize and less reliably measured) did not. If our operationalization of “pitch span” is perceived as friendliness, our finding is consistent with the direction of the (nonsignificant) result reported by [Groves and Benkí \(2006\)](#); pitch span may be more reliable than ratings of friendliness and so more likely to yield significant results. It is difficult to compare our results for pitch span with those of [Benkí et al. \(2013\)](#) because our measures are constructed in very different ways, and we predict outcome from pitch span, rather than describing the reverse.

Our results potentially inform measurements of propensity to participate. [Kennickell \(2012\)](#) found that ratings by field interviewers of the likelihood that a case would be ultimately interviewed in the Survey of Consumer Finances were too noisy to be useful. [Eckman, Sinibaldi, and Möntmann-Hertz \(2013\)](#) found that telephone interviewers have a modest ability to predict whether or not a sample member will ultimately be interviewed, but interviewer effects were large. In both these studies, the interviewers made the rating at the end of the contact, when considerably more information than “hello” was available. Because a high maximum pitch and the related pitch span of the sample member’s greeting predict participation, their potential as (relatively) external and reliable measures of propensity to participate could be explored. If recordings of the sample member’s “hello” could be analyzed at the speed required during field efforts, acoustic results could potentially be compared to or combined with other sources of information about the sample member’s propensity to participate, such as interviewers’ ratings, in responsive designs (e.g., [Groves](#)

8. Here are illustrative canonical and efficient introductions that begin with a disfluency, both from calls that end in a declination: “Uh good afternoon. I’m calling from University of Wisconsin uh for the Wisconsin Longitudinal Study for Mr. (FIRST AND LAST NAMES). Is he available?” and “Uh hello. May I please speak with (FIRST NAME)?”

and Heeringa 2006; Wagner, West, Kirgis, Lepkowski, Axinn, et al. 2012; Sinibaldi and Eckman 2015). Another potential application might be to train interviewers to recognize “large” and “small” greetings and to have a lower threshold for a “graceful exit” (as suggested by Schaeffer et al. 2013) from the latter type of call, in the hope of maximizing the chance of success on a later attempt.

We examined many acoustic properties of the interviewer’s greeting token: mean, minimum, and maximum pitch; pitch span; pitch pattern; duration; and latency. We operationalized acoustic reciprocity in several ways. Relationships were few, and some of those unexpected. One finding for interviewers suggests that a “large” greeting or “smile voice” might not be appropriate for a stranger calling: odds of participation are lower for interviewers in the top 30 percent of the distribution of maximum pitch. For acoustic reciprocity, we found that odds were lower when the interviewer mirrored a short greeting token. The relationship for latency is easier to explain: Odds of participation are higher for interviewers with the longest delay before speaking, which may provide an extra moment of processing or preparation.

Lexical reciprocity—the use of “hello” by the interviewer—had a positive effect on participation, but we cannot select among possible explanations for this (reciprocity, politeness, or fit to the expectations of older sample members). Our analysis of canonical introductions is consistent with a preference for a caller identifying themselves in their first turn (Schegloff 1979) and is similar to the observation by Campanelli, Sturgis, and Purdon (1997) in face-to-face interviews in a different population and to the judgment of experienced Dutch interviewers that it is important to “start by identifying yourself” (Snijkers, Hox, and De Leeuw 1999, pp. 192, 194).

Our findings might seem counter to suggestions that “conversational” introductions might be more effective than a script in recruiting survey participation (Houtkoop-Steenstra and van den Bergh 2002; also Morton-Williams 1993). However, the list of elements interviewers were required to include in that experiment (interviewer’s name, company name, research topic, phone number check, recipient selection, and number in the household—in any order) (Houtkoop-Steenstra and van den Bergh 2002, p. 207) is longer than the number of elements that our interviewers, using a “flexible introduction,” placed in the canonical turn. Moreover, that experiment did not include a manipulation check, so we do not know whether or how interviewers followed instructions, what interviewers actually included in the first turn, or what specific actions accounted for the observed effects.

Our study might imply that interviewers be trained and monitored on the content of a first turn modeled on the canonical turn examined here. However, other turn constructions not examined here may be at least as effective with this or other populations, so caution is called for in making such a recommendation. It is possible that the negative impact of an efficient introduction or the positive impact of the polite elements (minimal though they are) we observe is

specific to the cohort and study design represented by the WLS; a sample of younger people or a sample contacted on cell phones might have different sensibilities or prefer less polite formality. Still, for many studies, a household member of any age could be a gatekeeper, household informant, or selected sample member; moreover, caller identification must be accomplished in every population, and preferably before the sample member must ask “Who’s calling?”

Our design strengthens our predictions, but it has limitations. We can match pairs on estimated propensity to participate because we use data from a longitudinal study. But the overall response rate for the WLS is high enough that our small number of cases exhausts the pairs we could make with usable recordings, and so we cannot increase our sample size. The sample is homogeneous in race, origin, and age; most of our interviewers are considerably younger than the sample members; and these calls were made to landlines. Our sample members all have experience with the survey, most have received an advance letter, and interviewers could be fairly sure if the person who answered was not the sample member they sought. Because this was a panel study, the interviewer did not have to select a respondent from the household, and the placement of a selection procedure would have important consequences for the structure of the call opening; we could expect the opening sequence to be different in a cold call without a designated sample member (e.g., Maynard and Schaeffer 1997). All these features could affect which actions by the interviewer have consequences for participation.

However, our analysis of interviewers’ actions could facilitate experiments to design first turns for different target populations and emerging technologies. Study design (e.g., advance letters) and technology (e.g., caller identification) perform some aspects of “identification.” Although footing and social exchange theory provide ways of thinking about the interviewer’s first turn, that turn follows conventions for talk between strangers on the phone, conventions that continue to develop for cell phones and other modes of communication (Arminen and Leinonen 2006; Hutchby and Barnett 2005).

Supplementary Material

Supplementary materials are available online at http://www.oxfordjournals.org/our_journals/jssam/.

References

Arminen, I., and M. Leinonen (2006), “Mobile Phone Call Openings: Tailoring Answers to Personalized Summons,” *Discourse Studies*, 8, 339–368.

Banse, R., and K. R. Scherer (1996), “Acoustic Profile in Vocal Emotion Expression,” *Journal of Personality and Social Psychology*, 70, 614–636.

Benkí, J. R., J. Broome, F. Conrad, R. Groves, and F. Kreuter (2013), "Hello? Is Better Than Hello: Effects of Greeting Intonation on Participation in Survey Invitations," paper presented at the Annual Meeting of the American Association for Public Opinion Research, Boston, MA.

Benkí, J. R., J. Broome, F. G. Conrad, F. Kreuter, and R. M. Groves (2011), "Effects of Speech Rate, Pitch, and Pausing on Survey Participation Decisions," paper presented at the Annual Meeting of the American Association for Public Opinion Research, Phoenix, AZ.

Boersma, P., and D. Weenink (2012), "Praat: Doing Phonetics by Computer," Available at <http://www.fon.hum.uva.nl/praat/>.

Brown, P., and S. C. Levinson (1987), *Politeness: Some Universals of Language Use*, Cambridge: Cambridge University Press.

Campanelli, P., P. Sturgis, and S. Purdon (1997), *Can You Hear Me Knocking: An Investigation into the Impact of Interviewers on Survey Response Rates*, London: the Survey Methods Centre at SCPR, Social and Community Planning Research.

Chamberlain, G. (1980), "Analysis of Covariance with Qualitative Data," *The Review of Economic Studies*, 47, 225–238.

Conrad, F. G., J. Broome, J. R. Benkí, F. Kreuter, R. M. Groves, D. Vannette, and C. McClain (2013), "Interviewer Speech and The Success of Survey Invitations," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176, 191–210.

Couper, M. P., and R. M. Groves (2002), "Introductory Interactions in Telephone Surveys and Nonresponse," in *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*, eds. D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, and J. van der Zouwen, pp. 161–178, New York: Wiley.

De Leeuw, E., and J. Hox (1996), "The Effect of the Interviewer on the Decision to Cooperate in a Survey of the Elderly," in *International Perspectives on Nonresponse: Proceedings of the Sixth International Workshop on Household Survey Nonresponse, 25–27 October 1995*, Tutkimuksia Forskningsrapporter Research Reports, number 219, ed. Seppo Laaksonen, Helsinki: Statistics Finland, 46–52.

Dillman, D. A. (1978), *Mail and Telephone Surveys: The Total Design Method*, New York: John Wiley and Sons.

Dillman, D. A., J. D. Smyth, and L. M. Christian (2014), *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (4th ed.), Hoboken, NJ: Wiley.

Dykema, J., K. Diloreto, J. L. Price, E. White, and N. C. Schaeffer (2012), "ACASI Gender-of-Interviewer Voice Effects on Reports to Questions about Sensitive Behaviors Among Young Adults," *Public Opinion Quarterly*, 76, 311–325.

Eckman, S., J. Sinibaldi, and A. Möntmann-Hertz (2013), "Can Interviewers Effectively Rate the Likelihood of Cases to Cooperate?" *Public Opinion Quarterly*, 77, 561–573.

Goffman, E. (1979), "Footing," *Semiotica*, 25, 1–29.

Gouldner, A. W. (1960), "The Norm of Reciprocity: A Preliminary Statement," *American Sociological Review*, 25, 161–178.

Groves, R. M., and J. R. Benkí (2006), "300 Hello's: Acoustic Properties of Initial Respondent Greetings and Response Propensities in Telephone Surveys," paper presented at the 17th International Workshop on Household Survey Nonresponse, Omaha, NE.

Groves, R. M., and M. P. Couper (1996), "Contact-Level Influences on Cooperation in Face-to-Face Surveys," *Journal of Official Statistics*, 12, 63–83.

Groves, R. M., and S. G. Heeringa (2006), "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs," *Journal of the Royal Statistical Society, Series A*, 169, 439–457.

Groves, R. M., B. C. O'Hare, D. Gould-Smith, J. R. Benkí, and P. Maher (2008), "Telephone Interviewer Voice Characteristics and the Survey Participation Decision," in *Advances in Telephone Survey Methodology*, eds. J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, and R. L. Sangster, pp. 385–400, New Jersey: Wiley.

Hauser, R. M. (2005), "Survey Response in the Long Run: The Wisconsin Longitudinal Study," *Field Methods*, 17, 3–29.

Holtgraves, T., and J-N. Yang (1992), "Interpersonal Underpinnings of Request Strategies: General Principles and Differences Due to Culture and Gender," *Journal of Personality and Social Psychology*, 62, 246–256.

Houtkoop-Steenstra, H., and H. van den Bergh (2002), "Effects of Introductions in Large-Scale Telephone Survey Interviews," in *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*, eds. D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, and J. van der Zouwen, pp. 205–218, New York: Wiley.

Hutchby, I., and S. Barnett (2005), "Aspects of the Sequential Organization of Mobile Phone Conversation," *Discourse Studies*, 7, 147–171.

Kennickell, A. P. (2012), "What's The Chance? Interviewers' Expectations of Response in the 2010 SCF," Proceedings of the Survey Research Methods Section, The American Statistical Association.

Kockelman, P. (2004), "Stance and Subjectivity," *Journal of Linguistic Anthropology*, 14, 127–150.

Long, J. S. (1997), *Regression Models for Categorical and Limited Dependent Variables*, Thousand Oaks, CA: Sage.

Maynard, D. W., J. Freese, and N. C. Schaeffer (2010), "Calling for Participation: Requests, Blocking Moves, and Rational (Inter)action in Survey Introductions," *American Sociological Review*, 75, 791–814.

Maynard, D. W., and M. M. Hollander (2014), "Asking to Speak to Another: A Skill for the Telephone and Obtaining Survey Participation," *Research on Language and Social Interaction (ROLSI)*, 47, 28–48.

Maynard, D. W., and N. C. Schaeffer (1997), "Keeping the Gate: Declinations of the Request to Participate in a Telephone Survey Interview," *Sociological Methods and Research*, 26, 34–79.

McAleer, P., A. Todorov, and P. Belin (2014), "How Do You Say 'Hello'? Personality Impressions from Brief Novel Voices," *PLoS One*, 9, e90770.

McCulloch, S. K. (2012), "Effects of Acoustic Perception of Gender on Nonsampling Errors in Telephone Surveys," unpublished Ph.D. dissertation, Joint Program in Survey Methodology, University of Michigan–University of Maryland.

McCulloch, S. K., F. Kreuter, and S. Calvano (2010), "Interviewer Observed versus Reported Respondent Gender: Implications on Measurement Error," paper presented at the annual meeting of the American Association for Public Opinion Research, Chicago, IL.

Morton-Williams, J. (1993), *Interviewer Approaches*, Aldershot, UK: Dartmouth Publishing.

Nolen, J. A., and D. W. Maynard (2013), "Formulating the Request for Survey Participation in Relation to the Interactional Environment," *Discourse Studies*, 15, 205–227.

Oksenberg, L., and C. F. Cannell (1988), "Effects of Interviewer Vocal Characteristics on Nonresponse," in *Telephone Survey Methodology*, eds. R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, and J. Waksberg, pp. 257–272, New York: Wiley.

Oksenberg, L., L. Coleman, and C. F. Cannell (1986), "Interviewers' Voices and Refusal Rates in Telephone Surveys," *Public Opinion Quarterly*, 50, 97–111.

Pillet-Shore, D. (2012), "Greeting: Displaying Stance Through Prosodic Recipient Design," *Research on Language and Social Interaction*, 45, 375–398.

Purnell, T., W. Idsardi, and J. Baugh (1999), "Perceptual and Phonetic Experiments on American English Dialect Identification," *Journal of Language and Social Psychology*, 18, 10–30.

Schaeffer, N. C., D. Garbarski, J. Freese, and D. W. Maynard (2013), "An Interactional Model of the Call for Participation in the Survey Interview: Actions and Reactions in the Survey Recruitment Call," *Public Opinion Quarterly*, 77, 323–351.

Scharinger, M., P. J. Monahan, and W. J. Idsard (2011), "You had me at 'Hello': Rapid Extraction of Dialect Information from Spoken Words," *NeuroImage*, 56, 2329–2338.

Schegloff, E. A. (1979), "Identification and Recognition in Telephone Openings," in *Everyday Language: Studies in Ethnomethodology*, ed. G. Psathas, pp. 23–78, New York: Irvington.

— (1986), "The Routine as Achievement," *Human Studies*, 9, 111–151.

— (1998), "Reflections on Studying Prosody in Talk-in-Interaction," *Language and Speech*, 41, 235–263.

Scherer, K. R., R. Banse, H. G. Wallbott, and T. Goldbeck (1991), “Vocal Cues in Emotion Encoding and Decoding,” *Motivation and Emotion*, 15, 123–148.

Schweinberger, S. R., H. Kawarhara, A. P. Simpson, V. G. Skuk, and R. Zaske (2014), “Speaker Perception, WIREs,” *Cognitive Science*, 5, 15–25.

Sinibaldi, J., and S. Eckman (2015), “Using Call-Level Interviewer Observations to Improve Response Propensity Models,” *Public Opinion Quarterly*, 79, 76–93.

Snijkers, G., J. Hox, and E. D. de Leeuw (1999), “Interviewers’ Tactics for Fighting Survey Nonresponse,” *Journal of Official Statistics*, 15, 185–198.

Stephan, E., N. Liberman, and Y. Trope (2010), “Politeness and Psychological Distance: A Construal Level Perspective,” *Journal of Personality and Social Psychology*, 98, 268–280.

Sturgis, P., and P. Campanelli (1998), “The Scope for Reducing Refusals in Household Surveys: An Investigation Based on Transcripts of Tape-Recorded Doorstep Interactions,” *Journal of the Market Research Society*, 40, 121–139.

Tartter, V. C., and D. Braun (1994), “Hearing Smiles and Frowns in Normal and Whisper Registers,” *Journal of the Acoustical Society of America*, 96, 2101–2107.

Tourangeau, R. (2004), “Survey Research and Societal Change,” *Annual Review of Psychology*, 55, 775–801.

van der Vaart, W., Y. Ongena, A. Hoogendoom, and W. Dijkstra (2006), “Do Interviewers’ Voice Characteristics Influence Cooperation Rates in Telephone Surveys?” *International Journal of Public Opinion Research*, 18, 488–499.

Wagner, J., B. T. West, N. Kirgis, J. M. Lepkowski, W. G. Axinn, and S. K. Ndiaye (2012), “Use of Paradata in a Responsive Design Framework to Manage a Field Data Collection,” *Journal of Official Statistics*, 28, 477–499.