The Action Structure of Recruitment Calls and Its Analytic Implications
The Case of Disfluencies

Min, Bo Hee; Schaeffer, Nora Cate; Garbarski, Dana; Dykema, Jennifer

Link to publication in CBS Research Portal

The Action Structure of Recruitment Calls and Its Analytic Implications: The Case of Disfluencies

Bo Hee Min[1], Nora Cate Schaeffer[2], Dana Garbarski[3], and Jennifer Dykema[4]

We describe interviewers' actions in phone calls recruiting sample members. We illustrate (1) analytic challenges of studying how interviewers affect participation and (2) actions that undergird the variables in our models. We examine the impact of the interviewer's disfluencies on whether a sample member accepts or declines the request for an interview as a case study. Disfluencies are potentially important if they communicate the competence or humanity of the interviewer to the sample member in a way that affects the decision to participate. Using the Wisconsin Longitudinal Study, we find that although as they begin, calls that become declinations are similar to those that become acceptances, they soon take different paths. Considering all recruitment actions together, we find that the ratio of disfluencies to words does not predict acceptance of the request for an interview, although the disfluency ratio before the turning point – request to participate or a declination – of the call does. However, after controlling for the number of actions, the disfluency ratio no longer predicts participation. Instead, when we examine actions before and after the first turning point separately, we find that the number of actions has a positive relationship with participation before and a negative relationship after.

Key words: Participation; nonresponse; disfluencies; recruitment; survey introduction; interviewer-respondent interaction.

[1] Copenhagen Business School, Department of Management, Politics and Philosophy, Porcelænshaven 18B, 2000 Frederiksberg, Denmark. Email: bhm.mpp@cbs.dk
[2] University of Wisconsin-Madison, Department of Sociology, 1180 Observatory Drive Madison, WI 53706, U.S.A. Email: schaeffe@ssc.wisc.edu
[3] Loyola University Chicago, Department of Sociology, 440 Coffey Hall, 1032 W. Sheridan Rd. Chicago, IL 60660, U.S.A. Email: dgarbarski@luc.edu
[4] University of Wisconsin Survey Center, 475 No. Charter Street, Madison, WI 53706, U.S.A. Email: dykema@ssc.wisc.edu

1.   Introduction

Although survey interviews have been regularly conducted by phone for decades, we know surprising little about the sequence and structure of actions within the opening of these calls or about the implications of this structure for measurement and analysis. This article contributes to filling this gap in two ways. First, by providing a detailed description of the actions of the interviewer in calls to recruit a sample member, we show how the sequence of actions in calls that end in declination differs from that in calls that end in acceptance. Second, we present a case study that shows that taking the action structure of the call seriously affects conclusions. To do this, we examine whether disfluent speech – such as "um" and other fillers – by the interviewer predicts whether a sample member accepts or declines the request for an interview. Our case study explores the conclusion of an earlier study that the likelihood of acceptance of the request to participate was greatest when interviewers were moderately disfluent (Conrad et al. 2013). That conclusion suggested that disfluencies might be consequential because they communicate the competence or humanity of the interviewer to the sample member in a way that affects the decision to participate. Our analysis makes salient that disfluencies originate in an underlying structure of actions: the level of disfluencies by interviewers in a call depends on which actions are performed, the number of those actions, and the characteristic number of words in and level of disfluency of those actions.

To describe the structure of recruiting calls, we take advantage of an existing case-control design extracted from the Wisconsin Longitudinal Study (WLS) that compared two key outcomes of the initial contact with the sample member – declinations and acceptances. A conversation analysis of calls made for the WLS was the basis for an interactional model of the recruitment call (Schaeffer et al. 2013; Maynard et al. 2010). Grounding our analysis in this earlier work, we first describe the action structure of calls that end in acceptance and declination. We next motivate an interest in disfluencies and show how they are distributed over and located in the various actions by interviewers during recruitment. As a last step, we use the case-control design to predict acceptance of the request for survey participation from the disfluencies in the interviewer's speech and other variables with which those disfluencies are highly associated.

We find that which actions are performed, how many of them, and their typical levels of disfluency differ for calls that end in acceptance or declination. Some key actions of interviewers necessarily differ in calls that end in acceptance (e.g., talk about when to begin the interview) compared to those that end in refusals (e.g., responding to a refusal). Considering all recruitment actions together, we find that the ratio of disfluencies to words does not predict acceptance of the request for an interview, although the disfluency ratio before the turning point – request to participate or a declination – of the call does. However, after controlling for the number of actions, the disfluency ratio no longer predicts participation. Instead, when we examine the relationship between the number of actions and the odds of participation before and after the first turning point separately, we find that the number of actions has a positive relationship with the odds of participation before and a negative relationship after. In order to train interviewers to be successful in

recruiting sample members, it is important to be able to identify which features of which actions engage – or disengage – sample members.

2. Challenges in Studies of Interaction During Recruitment Calls

Studies about how interviewers influence the outcome of recruitment calls face substantial challenges. Advances in recording, transcribing, and coding interaction have allowed us to observe how closely the actual events during a recruitment call match our impressions (e.g., Dijkstra and Smit 2002; Maynard et al. 2002; Maynard and Schaeffer 1997; Schaeffer et al. 2013). For example, although we are rightly concerned about how best to train interviewers to address sample members' concerns (e.g., Groves and McGonagle 2001), sample members frequently exit without providing interviewers opportunities to use those skills (e.g., Sturgis and Campanelli 1998; Schaeffer et al. 2013). Similarly, the finding that householders who ask questions are more likely to participate (Groves and Couper 1996), can be refined to distinguish between questions that come before the request to participate (associated with a lower likelihood of participation) and questions placed after the request (a higher likelihood of participating) (Schaeffer et al. 2013). We know that many sample members stay on the phone call for only a few seconds, and that implies that we need to know what constitutes the most effective first turn for the interviewer, because it is the only talk that many sample members hear (Schaeffer et al. 2018).

Examining the impact of the interviewer's talk and actions during recruitment also raises technical issues of several kinds. First, because the sample member speaks first when they answer the phone, every action by the interviewer is plausibly influenced by the preceding actions of the sample member. Attributing a causal influence to any specific action by the interviewer requires strong study design. In the absence of an experiment, it may be ultimately unclear whether, for example, interviewers deliver more scripted descriptions of the study in cases that end in acceptance because such sample members are receptive when interviewers describe the study or because such scripted requests are persuasive (Schaeffer et al. 2013). Second, although we might have theoretical reasons to think that turns, actions, words, or some other feature of interaction, such as disfluencies of speech, are likely to be a critical influence on the sample member's decision, these are all very highly correlated in practice. Third, as the descriptions below show in detail, calls that end in acceptance look very different from calls that end in declinations. For example, in a large proportion of declinations, the sample member hangs up before the interviewer issues a request for participation, but the request is delivered in all but a handful of calls that end in acceptance (Schaeffer et al. 2013, and see detail below). If different sequences of actions lead to different outcomes, this also raises questions of measurement. For example, can a measure, such as the number of words in an interaction, be meaningfully compared for declinations and acceptances, when those words are produced in very different actions?

3. Interactional Model of the Recruitment Call

Our description of the actions that interviewers and sample members perform and the sequence of some of those actions is based on the interactional model of the recruitment call proposed in Schaeffer et al. (2013) (see also refinement in Schaeffer,

forthcoming). That earlier work identified the actions in the call but did not describe their relative frequency or how their frequency differed for calls with different outcomes. The recruitment encounter begins when the sample member comes to the phone and ends with a hang-up (for declinations) or when the interview begins (for acceptances).

Table 1 shows an example of a call that was initially a declination but ended in acceptance (one of a handful of within-in call conversions in our data), with the interviewer's recruitment actions labeled. The actions comprise roughly three phases with fluid boundaries: Identification, purpose of call, and recruitment. The identification phase includes the greeting, self and institutional identification, and request to speak to the sample member. As Table 1 suggests, these three actions regularly occupy the interviewer's first turn after the sample member answers (see also Schaeffer et al. 2018). The purpose of the call may be conveyed by the institutional identification in the call opening, but it is explicit in the second phase, which includes actions that verify the identity of the sample member and refer to the advance letter or the study. The recruitment phase includes the request for participation, attempts to persuade after any subsequent declinations ("follow-up to declination"), and statements that refer to the length of the interview (which at this location in the call are about scheduling the interview).

An analysis comparing calls that end in acceptance and declination must consider the overall structure of the calls. Although the calls begin similarly, there is a "turning point" at which their paths diverge. (See definitions in Table 2.) The interviewer's first request for participation, when it occurs, serves as the first turning point in the call, after which the outcome almost always quickly unfolds. When the interviewer is not able to deliver the request, the outcome is already unfolding, and the first turning point becomes the last interviewer action before the first declination, hang-up, or acceptance. The turning point is important analytically for at least two reasons: First, because many sample members exit during or quickly after the interviewer's first turn, it is possible that actions of the interviewer very early in the call have strong effects on some sample members (e.g., Schaeffer et al. 2018). Second, recruitment calls are most comparable before the first turning point, that is, during the phases of identification and explaining the purpose of the call (when those phases occur). After the first turning point, interviewers are either scheduling the interview (for acceptances) or attempting to persuade (after a refusal).

The analysis of actions, their relative frequency and sequence allow us to describe the differences in the actions that precede two key outcomes, acceptance and declination. We then apply this analysis to understanding how a feature of interviewers' behavior that has been of interest to other investigators – disfluency during the recruitment call (e.g., Van der Vaart et al. 2006; Conrad et al. 2013; Schaeffer et al. 2013; Schaeffer et al. 2018) – might influence the sample member's participation.

4. Disfluencies in Interviewers' Recruitment Actions: A Case Study

Disfluencies are non-lexical components of speech that take several forms, potentially including fillers (predominantly "um" and "uh"); the broken-off talk and repetitions that result from false starts and repairs ("ma- may I" or "I I am coming"); pauses; and

Table 1. Call opening illustrating interviewer's recruitment actions, WLS, call ending in acceptance.

Phase	Turn number	Interviewer's recruitment action	Actor	Talk
Identification	1		SM	Hello.
	2	Greeting	INT	Hello.
	2	Self-identification		My name is (FIRST NAME LAST NAME).
	2	Institutional identification		I'm calling from the University of Wisconsin Survey Center at the University of Wisconsin Madison.
	2	Request to speak to sample member		and I'm hoping to speak to Mr. (FIRST NAME LAST NAME)
	3		SM	Yes.
Purpose of call	6	Verification of sample member	INT	Is this the (FIRST NAME LAST NAME) who was enrolled at (NAME OF HIGH SCHOOL) High School in 1957?
	7		SM	Yes.
	8	Letter reference	INT	As you um may have you may recall from our recent letter,
	8	Study reference		we're doing a follow up study (inaudible) of our sample of people who were Wisconsin high school seniors in 1957,
	9		SM	Mhmm
Recruitment	10	Request to participate	INT	and we'd like to interview you now for this important study.
	10	Request to participate		Is this a good time for you?
	11		SM	If it doesn't take too long.
	12	Length-of-interview statement	INT	It does take about an hour.
	13		SM	Then no.
	16	Follow-up to declination	INT	We can it can be done in parts so we can just do as much of it as you like and then we can reschedule um to call you back another time that's better for you.
	17		SM	Yeah. What w- what if we can do like fifteen minutes. Would be alright, I guess, but otherwise I've got things I gotta do.

Note: SM is the sample member; INT is the interviewer. Within a turn, each action is shown on a separate line. Punctuation has been added and minor actions (e.g., exclamations) deleted for readability.

Table 2. *Concepts and definitions.*

Type of concept	Concept	Definition and coding
Structure of call	Interviewer's recruitment actions	A unit of talk in the opening of the recruitment call that accomplishes a specific interactional task. Actions were identified as regular components of the call in a conversation analysis. A turn may include more than one action. See examples and labels in Table 1. Source: Maynard, Freese, and Schaeffer 2010; Schaeffer et al. 2013. (See also concept of "moves" in Conrad et al. 2013.)
	First turning point	For acceptances, the first turning point is the first request for participation or the last interviewer action before an acceptance. For declinations, the first turning point is the last interviewer action before the first (blocking) declination, hang up, or acceptance, whichever came first. Sources: Schaeffer et al. 2013; see also Conrad et al. 2013.
	Congruent/ incongruent actions	Actions that "follow up" declinations are common in calls that end in declination and uncommon in calls that end in acceptance; the follow-up to declination is "congruent" with the declination path. Similarly, talk about the length of the interview and how to administer it in parts are common in calls that end in acceptance and uncommon in calls that end in declination; the "statement of length of interview" is congruent with the acceptance path.
Disfluencies	Disfluency	Irregularities in speech that intrude into the smooth production of talk. We examine three possible disfluencies: fillers, broken-off talk that results from false starts and repairs, and nonpropositional elements. Source: Bortfeld et al. 2001; Schober et al. 2012.
	Fillers	A set of non-lexical tokens. Our data include "Uh Um Ah Hmm Mmm Eh Aw Er Nn Num," and we standardized their spelling to facilitate accuracy of text coding. Source: Bortfeld et al. 2001.
	Nonpropositional elements	A cover term for words that are used as "discourse markers", for example, around fillers at the beginning of a turn (e.g., "and um and"), or as "acknowledgments" (e.g., "okay," "okay," speaker changes. The most frequent nonpropositional elements in our corpus are "and," "okay," "alright," "so," "well," and "but." Most of the remaining were synonyms for these words and appeared four or fewer times (e.g., "excellent" instead of "okay"). We counted nonpropositional elements immediately preceding one or more fillers (that immediately precede an action). Source: Adapted from Bortfeld et al. 2001.
	Broken-off talk	Broken-off talk occurs when speaker stops talking before completing a unit of talk, usually an action. Broken-off talk was indicated by trailing "-" in the transcripts. Broken-off talk was usually followed by a restart (e.g., "I-I was"). We used the length of a pause at speaker change to distinguish broken-off talk from interruptions. If talk overlapped at the transition, we considered the break-off to be due to an interruption.Source: Our broken-off talk is similar to one component of Bortfeld et al.'s (2001) "restart." Due to our limited resources, we did not examine all restarts, which can be quite complex.

"nonpropositional elements," which include discourse markers ("but" or "well") and acknowledgment tokens ("okay"), that precede or are embedded with fillers (Bortfeld et al. 2001). See Table 2 for definitions used here.

4.1.   Interviewers' Disfluencies: Theoretical Issues and Prior Research

In studies of survey interviews, most attention has been given to disfluencies of respondents during the interview itself: Disfluent respondents may be treated as having comprehension problems (Schaeffer and Maynard 2002; see also Schober and Bloom 2004), and respondents' disfluencies may indicate that an answer is less accurate or reliable (Draisma and Dijkstra 2004; Draisma et al. 2005; Schaeffer and Dykema 2011; Garbarski et al. 2011; Schober et al. 2012; Smith and Clark 1993; Mathiowetz 1999).

The impact of the interviewer's disfluencies on recruitment may depend on the perceptions of sample members (e.g., Van der Vaart et al. 2006). Sample members may ignore disfluencies; or disfluencies may affect whether the interviewer is perceived as comfortable, confused, honest, anxious, and so forth (e.g., Christenfeld 1995; Fox Tree 2002, 2007). Disfluencies may perform other tasks that are informative: Listeners may hear a disfluency as signaling that the next item mentioned may be new (Arnold et al. 2004; Arnold et al. 2007; Barr and Syfeddinipur 2010) or that a repair is forthcoming (Brennan and Schober 2001). Disfluencies may also separate "intonation units" (Clark and Fox Tree 2002) in a way that may serve as audible "punctuation" and so make speech easier to understand. "Uh(m)" may serve to delay dispreferred acts or the purpose of a call (Schegloff 2010).

In studies of interviewers' success in recruiting sample members, disfluencies, as defined here, have been studied less than other acoustic and behavioral qualities of the interviewer's speech (e.g., Groves et al. 2008). For an interviewer, being disfluent may (or may not) be associated with whether the interviewer is successful at recruiting sample members (Schaeffer et al. 2013; Schaeffer et al. 2018; Conrad et al. 2013; Oksenberg and Cannell 1988; Van der Vaart et al. 2006; Sharf and Lehman 1984). Schaeffer et al. (2013) found higher odds of participation when disfluencies were present; however, they also predicted higher odds of participation if the interviewer's mention of the advance letter or description of the study followed an available (optional) script, a practice that reduced disfluencies (results not shown). Their subsequent analysis of the interviewer's first turn indicated that the odds of participation were lower ($p < 0.10$) if that turn began with a filler (Schaeffer et al. 2018), although few first turns began in this way. In an analysis with multiple samples and a different design, Conrad et al. (2013, 201) found that participation had a curvilinear relationship with the interviewer's filler rate (fillers per 100 words): The proportion of householders who agreed to participate was lowest for interviewers with the lowest (0) or highest filler rate. Their interpretation was that interviewers with no disfluencies may sound robotic, and those with too many disfluencies may sound incompetent.

4.2.   Interviewers' Disfluencies and Their Locations

As have other researchers, we observe that interviewers' disfluencies regularly occur in three locations: At the beginning of a turn, at the beginning of an action within a turn, and within an action (Boomer 1965; Shriberg 1996; Clark and Fox Tree 2002, 95). Excerpt 1

Excerpt 1.   Interviewer's canonical introduction, showing fillers before next action, call that ends in declination, WLS, punctuation and capitalization added.

Line	Turn	Action	Transcript	Disfluency and location
1	4	Greeting	Hi, sir.	
2	4		uh	Filler before next action
3	4	Self-identification	My name's (FF) (LL).	
4	4	Institutional identification	I'm calling from the University of Wisconsin in Madison for the Wisconsin Longitudinal Study.	
5	4		um	Filler before next action
6	4	Letter reference	Did you happen to get our letter in the mail recently?	

shows how disfluencies are located within the actions of the call opening, which suggests ways they might function.

Excerpt 1 begins after the interviewer delivered an "efficient" introduction (which begins by confirming that he is speaking with the sample member instead of with self-identification, not shown). Excerpt 1 begins with turn 4, the interviewer's second turn, in which he adapts the "canonical" introduction (one that begins with greeting and self-identification – provided on his screen) (Schaeffer et al. 2018) and adds a reference to the advance, because WLS interviewers were authorized to treat the scripted introduction as "flexible" (Morton-Williams 1993; Houtkoop-Steenstra and Van den Berg 2002). He inserts the fillers "uh" and "um" in lines 2 and 5 before his "self-identification" and "letter reference" actions, so that the disfluencies reinforce the meaningful units within the interviewer's stream of talk.

In Excerpt 2, the request to speak to the sample member at line 8 begins with a discourse marker ("and"), followed by a filler ("uh"), and then broken-off talk ("we were wonder-"), followed by another filler ("uh") and a restart or repair ("we were wondering"). The midstream embedded disfluencies ("we were wonder- uh") do not mark transitions of speaker or action the way the "uh" at line 4 or the initial "and uh" in line 8 do.

These excerpts suggest that a disfluency in any location may indicate that the speaker is planning speech, retrieving words, or undertaking a repair. Disfluencies that occur midstream during an action or turn (e.g., line 8 in excerpt 2) may be distinct, either in their origins or in how they are perceived by listeners. These midstream or embedded disfluencies do not perform the turn-taking or transitional work performed by disfluencies that begin an action or turn; they may communicate that the speaker is searching for what to say or how to say it and so be more consequential.

4.3.   Hypotheses: Interviewers' Actions, Disfluencies, and Participation

Prior investigations of the impact of the interviewer's disfluencies on participation have not accounted for how they are located within actions. Our hypotheses examine the

Excerpt 2. Interviewer's second introduction with sample member, illustrating discourse markers and fillers, call that ends in declination, WLS, punctuation and capitalization added, SM = sample member, INT = interviewer.

Line	Turn	Action	Actor	Transcript	Disfluency
1	3		SM	You spoke so fast I couldn't understand who this was.	
2	4	Change-of-state Token	INT	Oh,	Discourse marker at beginning of turn
3	4	Apology		I'm sorry.	
4	4			uh	Filler before next action
5	4	Self-identification		My name's (FF) (L).	
6	4	Institutional identification		I'm calling from the University of Wisconsin Madison for the W L S study the	
7	5		SM	Okay.	
8	6	Request to speak to sample member	INT	And uh we were wonder- uh we were wondering if we could speak to dzhu- (FF) (LLL).	Discourse marker, filler, broken-off talk, filler, restart at beginning of turn
9	7		SM	This is she.	

components of a sample member's experience of disfluencies: Disfluencies occur in actions – which can vary in frequency, length, and fluency – at various locations in the call.

We first compare the action structure of calls that end in declination and acceptance. For example, it is possible that some calls with no disfluencies end in refusal because the householder hangs up before the interviewer has much opportunity to talk (see discussion in Sturgis and Campanelli 1998) and thus to be disfluent. Similarly, the level of disfluency could be high if interviewers become flustered and increasingly disfluent when trying to persuade very resistant sample members. The model of the call, the literature briefly reviewed above, and our observations of disfluencies lead to our first prediction, that the action structure of the call opening will be similar for calls with that end in acceptances and declinations but diverge after that.

We then turn to the fluency of actions. The varied actions in the call make different demands on the interviewer. The identification phase is familiar and well-rehearsed and likely to be delivered fluently. Once the identification phase is complete, the behavior of the sample member becomes less predictable, and sometimes hostile, and the interviewer must plan and execute actions quickly. When a sample member declines, the interviewer may perceive the stakes as higher; the interviewer's actions following a declination take place in an uncertain environment, and the interviewer may be more, and differently, disfluent as a result. This leads to our second prediction, that the most routine actions, in the identification phase of the call, will have low rates of disfluency that are similar

for calls with each outcome, and most disfluencies that do occur will be placed at the beginning of actions. Later in the call, and particularly after a declination, the uncertainty of the interactional environment will lead to more and more varied disfluencies, including midstream disfluencies of broken-off talk and midstream fillers. This leads to our third prediction, that actions that follow up declinations, an action congruent with declinations, will be less fluent than talk about the length of interview (used to schedule an interview after an acceptance), an action congruent with acceptances.

If we observe these differences in action structure and in the fluency of various actions, we can appreciate in a different way the complexity of estimating the impact of the interviewer's disfluency on participation: For example, if interviewers are most disfluent when trying to persuade reluctant sample members, and these disfluencies do not appear for accommodating sample members, then we might suspect that these disfluencies result, at least in part, from the sample member's resistance (or propensity to participate) and so might be effects of a decision the sample member has already made. So, to complement our descriptive analysis, we then take advantage of our case-control design (see below) to predict participation from disfluencies. Because the structure of calls that end in declination and acceptance are most comparable before the first turning point in the call (results below), we examine the impact of our measures before and after that point. Because calls move quickly to their outcome after the first turning point, we might predict that disfluencies before the first turning point make a poor impression and reduce the likelihood of participation. We predict that calls with more disfluencies after the first turning point – for example, because interviewers get flustered attempting to persuade resistant sample members – will be more likely to end in declination. However, we expect that the number of disfluencies, words, and actions are interdependent in ways that makes it difficult to assess the impact of each, and we address this by examining the impact of disfluencies net of the number of actions.

5.   Methods: Data, Variables, and Analysis

Our analysis requires detailed transcripts and some ability to estimate the impact – or, at least, predictive strength – of features of interaction on participation. Designing an experiment to address this topic in a production context would present substantial obstacles. So, we address our research questions using a recent case-control design constructed from the Wisconsin Longitudinal Study (WLS). See details of study design in Schaeffer et al. (2013). However, because we use matched pairs, our declinations and acceptances cannot be combined to estimate characteristics of the WLS. (For example, our sample, by design, has a "response rate" of 50%, but the response rate for the WLS is much higher.) Considered separately, our declinations and acceptances each constitute a collection or corpus, rather than a probability sample from a specific population, although some of our tests treat them as independent samples.

5.1.   Sample

Our analysis uses digital records of phone contacts from the 2004 round of the WLS, which interviewed 80% of surviving panel members. The WLS began with a one-third

sample of 1957 Wisconsin high school graduates and followed up in 1964 (mail to parents), 1975 (telephone), 1992 (telephone and mail), and 2004 (telephone and mail). (See Hauser 2005.) The case-control study selected 257 pairs of cases (the maximum number of pairs that could be made). One member of each pair declined to be interviewed in their first contact with the interviewer (declination), and the other member of the pair accepted on the first contact with an interviewer (acceptance). Pair members matched exactly on sex and past participation in the WLS and as closely as possible on estimated propensity to participate. The model estimating the propensity to participate included education, high school class rank, high school cognitive assessments, self-reported health, sex, and past participation (See Appendix, Section 8, for additional details.). To the extent that the pairs are successfully matched on propensity to participate, differences in outcome should be largely due to the behavior of the interviewer. We recognize, however, that the matching of the pairs is subject to measurement and other errors, and we modulate our claims of causality accordingly.

5.2. Analysis of Actions

Actions (listed in the example in Table 1) and their features were identified in an extended conversation analysis of the call opening. The interactional model of the recruitment call summarizes this analysis and the reliability of coding is described elsewhere (Maynard et al. 2010; Schaeffer et al. 2013; Schaeffer et al. 2018; Schaeffer, forthcoming). The definitions of the concepts in the present analysis are summarized in Table 2 .

5.3. Variables: Measuring Disfluencies

Disfluencies have been operationalized in many ways. Our definitions (see Table 2 and Appendix) drew heavily on the concepts and operational rules described by Bortfeld et al. (2001, 131–132) because of their relevance and completeness. We developed computer code to identify and count disfluencies in transcripts that had been standardized in preparation. The summary statistics we discuss are described in Table 3, and the Appendix gives details of underlying rules for counting disfluencies.

5.4. Analysis

Our analysis has two parts. First, our descriptive analysis examines the components of exposure to disfluencies, which comes by way of the specific actions the interviewer performs and the frequency, length, and fluency of those actions: We first describe the action structure of the calls that end in acceptance and declination, and then we describe the fluency of those actions. This decomposed description is suitable for our corpus of acceptance and declination calls. This decomposition is useful because other populations of sample members and interviewers and other study designs could give rise to a different distribution of actions or different levels of disfluency in those actions. Results from our approach might be generalizable if these actions appear in other studies; for example, if interviewers self-identify in similar ways across a variety of populations, the effects of self-identification might then be expected to be similar. In the descriptive analysis, when

Table 3. Calculation of summary measures.

Table	Label in table	Unit	Summary statistic	Calculation
5	Mean turn number of the action, all actions of a given type	Actions of given type	Mean of turn numbers of all actions of given type	Within each call, each action of a given type was assigned its turn number. (The sample member's answer to summon is turn 1.) The mean turn number of all actions of a given type was calculated. A call may have more than one action of a given type.
5	Modal turn number of the action, all actions of a given type	Actions of given type	Mode of turn numbers of all actions of given type	Within each call, each action of a given type was assigned its turn number. (The sample member's answer to summon is turn 1.) The most common value of the turn numbers of all actions of the type is calculated. A call may have more than one action of a given type.
5	Mean of number of actions per call, for calls with the action	Calls with action of given type	Sample-level mean of number of actions of given type per call	Within each call, the number of actions of given type is counted. The mean number of actions of a given type is calculated. If a call has no action of a given type, it is excluded from the calculation.
5	Mean of mean number of words per action, for calls with the action	Calls with action of given type	Sample-level mean of call-level mean number of words in each action (excluding fillers and broken-off talks), for calls with at least one instance of given action.	Within each recruitment action, the number of words in each action (excluding fillers and broken-off talk) is counted. Within each call with at least one action of a given type, the mean of the number of words in actions of a given type is calculated. Then the sample-level mean of call-level means is calculated. If a call has no action of a given type, it is excluded from the calculation.

Table 3. Continued.

Table	Label in table	Unit	Summary statistic	Calculation
6	Initial: mean of proportion of actions that begin with any disfluency	Calls with action of given type	Sample-level mean of call-level proportion of actions of given type with initial disfluency	Within each action, a dummy variable indicates if there is a initial disfluency. Within each call, the proportion of actions of a given type with a initial disfluency is calculated. Then a sample-level mean of these proportions is calculated. If a call has no action of a given type, it is excluded from the calculation.
6	Midstream: mean of proportion of actions with any midstream filler or broken-off talk	Calls with action of given type	Sample-level mean of call-level proportion of actions of given type with midstream disfluency	Within each action, a dummy variable indicates if there is a midstream disfluency. Within each call, the proportion of actions of a given type with midstream disfluency is calculated. Then a sample-level mean of these proportions is calculated. If a call has no action of a given type, it is excluded from the calculation.
6	Mean of proportion of actions with any disfluency in any location	Calls with action of given type	Sample-level mean of call-level proportion of actions of given type with disfluency at any location	Within each action, a dummy variable indicates if there is a disfluency at any location. Within each call, the proportion of actions of a given type with any disfluency is calculated. Then a sample-level mean of these proportions is calculated. If a call has no action of a given type, it is excluded from the calculation.

Table 3. Continued.

Table	Label in table	Unit	Summary statistic	Calculation
6	Mean of mean number of disfluencies in any location	Calls with action of given type	Sample-level mean of call-level mean of number of disfluencies in actions of a given type in the call	Within each action, the number of disfluencies (fillers, broken-off talk, and discourse markers) is calculated. Within each call, the mean number of disfluencies in actions of a given type is calculated. Then a sample-level mean of these means is calculated for calls with actions of a given type. If a call has no action of a given type, it is excluded from the calculation.
7	Number of disfluencies	Calls	Total count of disfluencies in the action or call	Within each of the recruitment actions, fillers at any location in the action, broken-off talk, and discourse marker are counted. Within each call, these numbers are summed across the recruitment actions.
7	Number of words/100	Calls	Total count of words in recruitment actions or calls, excluding disfluencies, divided by 100	Within each of the recruitment actions, words (excluding fillers at any location in the action, broken-off talk, and discourse marker) are counted. Within each call, the numbers words across recruitment actions is summed and divided by 100.
7	Number of actions	Calls	Total count of recruitment actions or actions of a given type in the call	Within each call, the number of recruitment actions (see Table 1) or actions of a given type is counted.
7	Disfluency ratio	Calls	(number of disfluencies/ (number of words/100))	Within each of the recruitment actions, the number of disfluencies (fillers at any location in action, broken-off talk, and discourse markers) and the number of words excluding disfluencies is counted. Within each call, the total number of disfluencies and the total number of words excluding disfluencies are counted. The ratio of the total number of disfluencies to the (total number of words/100) is calculated.

we provide t-tests for differences in means or proportions, we treat the acceptances and declinations as independent samples from the WLS.

In the second part of our analysis, we use the case-control design (conditional logit analyses, clogit in Stata) to predict participation as the dependent variable (see Schaeffer et al. 2013; Schaeffer et al. 2018). The following likelihood function for clogit with groups (that is, pairs of observations) was used:

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

where

- 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 indicates that there may be more than one regressor in the model
- i is the group identifier
- ij, where $j \in \{1,2\}$, is the jth observation of the ith group
- $I_1 = \{i \,|\, y_{i1} + y_{i2} = 1\}$
- x_{ij} is the row of covariates associated with the jth observation of the ith group
- $I(j = 2)$ is the indicator function for $j = 2$

The outer summation is over all pairs in which the pair's responses contain one 0 (declination) and one 1 (acceptance). The inner summation is over the single observation within the pair in which the response is 1. (The likelihood function minimized by clogit is described on the Stata clogit page (http://www.stata.com/manuals14/rclogit.pdf). This section references several other sources, including Chamberlain (1980), which is the basis for the likelihood function above (Mark Banghart, personal communication).

Conditional logit is similar to a fixed-effect logit in which the matching characteristics (see above) 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 logit 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 discuss relationships that are significant with the relatively generous $\alpha = 0.10$ but note when results are significant by conventional standards ($\alpha = 0.05$).

6. Results

We provide descriptions that have been absent from the literature to date: (1) a detailed picture of the overall action structure of the recruitment call, and (2) variation in disfluencies by location of disfluency (at beginning of action or midstream within an action), and by type of action. We then use information about actions and disfluencies to predict acceptance of the request for participation.

6.1.   Interviewers' Actions in Recruitment Calls: Overall Structure

Table 4 and Table 5 describe interviewers' actions. Table 4 gives counts of actions (Panel A) and of calls with actions (Panel B and Panel C). These counts are descriptive in themselves and also document the number of units on which the summary statistics in Table 5 and Table 6 are based. Table 5 describes other features of actions: the turns in which they are located (Panel A) and their number and length (in words) (Panel B), for calls with each outcome.

Results in both tables confirm that, overall, the identification phase is similar for declination and acceptances calls. In Table 4, Panel B demonstrates that the single significant difference (in issuing a greeting) is substantively small. In Table 5, Panel A reinforces the similarity of the actions in the identification phase: The mean (column (a)) and modal turn numbers (column (b)) of the actions are similar or identical across actions and outcomes.

After identification some declining sample members have hung up, and the trajectories of remaining calls destined for acceptance or declination increasingly diverge: the number of calls with each action after the identification phase is significantly different for declinations and acceptances (Table 4, Panel B). More calls that end in acceptance (compared to calls that end in declinations) have each of the actions in the "purpose of call" phase: the interviewer's verification that they have reached the sample member, questions about the advance letter, and descriptions of the study (Table 4, Panel B). Differences in the sequential position of some actions also begin to appear after the identification phase (Table 5, Panel A): For example, the modal turn number for the letter reference and request to participate is one turn later for calls that end in acceptance than for declinations. For the "purpose of call" actions, the mean turn numbers are higher than the mode, as the call structure becomes less conventional.

A critical difference in the action structure of calls with the two outcomes is that the request to participate occurs in almost all acceptances, but in fewer than half of declinations (Table 4, Panel B). Of the 257 declinations, 15 have hung up and another 141 have declined and hung up before the interviewer can issue a request. In two cases the first request came after the first declination (detail not shown). This massive and selective exodus of sample members very early in the call means that many who decline have almost no exposure to the interviewer – they are exposed to few interviewer actions, disfluencies, or attempts at persuasion. In Table 4, column (d) in Panel C clarifies what Panel B suggests – that interaction after the first turning point is dominated by the congruent actions: Follow-up actions in declinations (in 175 of 257 calls) and talk about the length of the interview in acceptances (in 146 of the 257 calls).

Thus far, our results suggest that exposure to disfluencies might differ for declination and acceptance calls because the outcomes are preceded by different actions. Panel B in Table 5 indicates that the different frequency (column (c)) and length in words (column (d)) of various actions could also contribute. Once again, calls with both outcomes are similar in the identification phase: Comparing declinations and acceptances, the mean number of actions is similar for the two outcomes (with a single small difference), as is the mean number of words per action. In the "purpose of call" phase, the mean number of references to the advance letter is greater and the mean number of actions discussing the

Table 4. Number of recruitment actions and number of calls with each type of recruitment action by the interviewer, by location and call outcome.

		Number of actions		Number of calls in which action occurs...						
		Panel A		Panel B			Panel C			
		(a) Actions of given type		(b) Anywhere in the call			(c) Up to and including first turning point		(d) After first turning point	
Phase	Recruitment action	Declination	Acceptance	Declination	Acceptance	p	Declination	Acceptance	Declination	Acceptance
Identification	Greeting	434	430	257	253	**	256	253	11	3
	Self-identification	246	254	236	234		236	234	1	2
	Institutional identification	309	309	256	255		255	255	7	4
	Request to speak to sample member	230	236	208	209		207	208	2	2
Purpose of call	Verification of sample member	80	175	75	168	***	72	128	3	40
	Letter reference	205	343	157	233	***	147	224	17	24
	Study reference	247	316	128	200	***	117	182	31	44

Table 4. Continued.

Phase	Recruitment action	Number of actions		Number of calls in which action occurs...						
		Panel A		Panel B			Panel C			
		(a) Actions of given type		(b) Anywhere in the call			(c) Up to and including first turning point		(d) After first turning point	
		Declination	Acceptance	Declination	Acceptance	p	Declination	Acceptance	Declination	Acceptance
Recruitment and persuasion	Request to participate	129	394	101	253	***	99	253	9	69
	Follow-up to declination	928	26	175	10	***	0	0	175	10
	Length-of-interview statement	51	285	32	146	***	2	1	30	146
Total number of actions or calls		2,859	2,768	257	257		257	257	257	257

Note: Cells in Panel A are number of actions in calls of the type; cells in Panels B and C are number of calls in which an action occurred at least once. Cells for actions after the turning point that are "incongruent" with the outcome are in gray font; in addition, in Panel C, an action is shown in black font in the "up to or including" or "after" panel if the action occurs in at least 9 calls for both acceptances and declinations.

Base analytic sample is matched pairs: 257 acceptances and 257 declinations. Interaction begins when sample member answers or is brought to the phone and ends when the sample member hangs up (declination) or the interview begins (acceptance). For acceptances, the first turning point is the first request for participation or the last interviewer action before an acceptance. For declinations, the first turning point is the last interviewer action before the first (blocking) declination, hang up, or acceptance, whichever came first.

Table 5. Selected characteristics of calls with recruitment actions, by call outcome.

| Phase | Recruitment Action | Panel A Cells based on all actions of a given type | | | | Panel B Cells based on all actions of a given type | | | | | |
| | | (a) Mean turn number of the action, all actions of a given type | | (d) Mean of mean number of words per action, for calls with the action | | (d) Mean of mean number of words per action, for calls with the action | | | (d) Mean of mean number of words per action, for calls with the action | | |
		Declination	Acceptance	Declination	Acceptance	Declination	Acceptance	p	Declination	Acceptance	p
Identification	Greeting	1.52	1.54	1	1	1.69	1.70		2.19	2.24	
	Self-identification	1.77	1.79	2	2	1.04	1.09	*	5.44	5.37	
	Institutional identification	2.01	2.01	2	2	1.21	1.21		11.93	11.92	
	Request to speak to sample member	1.21	1.33	1	1	1.11	1.13		6.64	6.76	
Purpose of call	Verification of sample member	2.69	3.73	2	2	1.07	1.04		17.72	18.60	
	Letter reference	3.33	3.67	2	3	1.31	1.47	**	11.25	10.63	
	Study reference	4.98	5.07	3	3	1.93	1.58	*	23.71	21.29	*
Recruitment and persuasion	Request to participate	3.88	4.90	3	4	1.28	1.56	***	18.37	17.73	
	Follow-up to declination	6.90	5.65	4	4	5.30	2.60	***	14.12	21.77	*
	Length-of-interview statement	6.69	6.49	7	5	1.59	1.95	**	28.16	28.91	*

Note: In Panel A, statistics are calculated for actions of a given type; multiple actions of a given type within a call are all included; Panel A in Table 4 shows the number of actions of a given type. In Panel B, unit is calls in which an action of a given type occurred; Panel B in Table 4 shows the number of calls with an action of a given type. Tests in Panel B compare the mean for the two call outcomes. Because a call may have more than one action of a given type, means for number of words are calculated for each call in which an action occurs; the mean (of the means) for the calls that included a given action is reported in the table.

Cells for actions that are "incongruent" with the outcome are in gray font. Base analytic sample is matched pairs: 257 acceptances and 257 declinations. Interaction begins when sample member answers or is brought to the phone and ends when the sample member hangs up (declination) or the interview begins (acceptance).

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

study smaller for acceptance than declination calls. The recruitment phase is substantially different for the two outcomes: Column (c) shows that the mean number of requests to participate is larger for acceptance than declination calls, and thereafter the calls become even more difficult to compare. For example, in column (c) the mean number of "follow-up to declination" actions in declination calls is 5.3, and the mean number of actions about the length of interview in acceptance calls is 1.95. As shown in column (d), however, the former action is shorter on average (mean average number of words = 14.1) than are actions about the length of interview in acceptance calls (28.9). Thus, the relative frequency and, to a lesser extent, the relative length of different actions provide different opportunities for exposure to the interviewer's disfluency.

Taken together, the results in Table 4 and Table 5 support our first hypothesis: Calls with each outcome begin similarly but diverge sharply after the first turning point. We also find that a substantial number of sample members who decline have extremely short calls with few actions by the interviewer, and a substantial number have longer calls with multiple follow-up attempts by the interviewer. The actions that occur and their frequency and length vary for calls with different outcomes.

6.2. Disfluencies in Interviewers' Actions

Table 6 summarizes features of disfluencies in the recruitment actions by call outcome. (Table 3 gives details of calculations.) We examined many measures of disfluency (some of them overlapping) to understand how they differed and select among them. We present summary statistics and test the difference between acceptances and declinations. As part of our analytic approach that distinguishes the presence of an action and its features – a sort of decomposition strategy – we focus here on calls in which the action occurred in order to characterize disfluencies in various actions when they occur. Panel A of Table 6 presents the mean (across all calls with an action) of the proportion of actions that begin with a disfluency (initial disfluencies) (column (a)) or include a midstream disfluency or broken-off talk (midstream disfluencies) (column (b)). Panel B summarizes across components of disfluency, so that columns (c) and (d) of Panel B are different summaries of the information in columns (a) and (b).

In the identification phase, there is a single notable difference between declination and acceptance calls: Although a disfluency at the beginning of a greeting is rare, it appears more often in declinations (as reported in Schaeffer et al. 2018). As predicted, compared to actions later in the call, the actions in the identification phase are relatively fluent. An exception is the presence of initial disfluencies for self-identification, but the frequency is not significantly different for declinations and acceptances (Panel A and Panel B). It is plausible that an initial disfluency for self-identification simply separates it from a preceding greeting in the same turn (as in Excerpt 1, line 4). In the identification phase, midstream disfluencies are relatively frequent for "institutional identification," an action that allows interviewers to choose components of the identification, but there are no significant differences between declinations and acceptances.

In the "purpose of the call" and "recruitment and persuasion" phases, initial and midstream disfluencies each appear to be similarly frequent for both call outcomes. However, the level of disfluency appears higher in declination than acceptance calls for

Table 6. *Disfluencies in recruitment actions, for calls with actions of a given type, by call outcome.*

| | | Panel A. Components of disfluency | | | | | | | | | Panel B. Summary measures | | |
| | | (a) Initial: mean of proportion of actions that begin with any disfluency | | | (b) Midstream: mean of proportion of actions with any midstream filler or broken-off talk | | | (c) Mean of proportion of actions with any disfluency in any location | | | (d) Mean of mean number of disfluencies in any location | | |
Phase	Recruitment Action	Declinations	Acceptances	p	Declinations	Acceptances	p	Declinations	Acceptances	p	Declinations	Acceptances	p
Identification	Greeting	0.08	0.03		0.01	0.02		0.08	0.04		0.08	0.04	
	Self-identification	0.16	0.20	***	0.02	0.03		0.18	0.23	**	0.19	0.24	**
	Institutional identification	0.05	0.06		0.11	0.13		0.14	0.18		0.21	0.24	
	Request to speak to sample member	0.08	0.13	*	0.08	0.08		0.14	0.19		0.19	0.28	
Purpose of call	Verification of sample member	0.38	0.43		0.27	0.21		0.51	0.57		0.77	0.95	
	Letter reference	0.30	0.32		0.17	0.14		0.39	0.38		0.61	0.63	
	Study reference	0.23	0.21		0.32	0.26		0.46	0.39		1.08	0.75	**

Table 6. Continued.

		Panel A. Components of disfluency									Panel B. Summary measures		
		(a) Initial: mean of proportion of actions that begin with any disfluency			(b) Midstream: mean of proportion of actions with any midstream filler or broken-off talk			(c) Mean of proportion of actions with any disfluency in any location			(d) Mean of mean number of disfluencies in any location		
Phase	Recruitment Action	Declinations	Acceptances	p	Declinations	Acceptances	p	Declinations	Acceptances	p	Declinations	Acceptances	p
Recruitment and persuasion	Request to participate	0.37	0.28	*	0.27	0.23		0.52	0.42	*	0.93	0.74	
	Follow-up to declination	0.22	0.15		0.21	0.49	*	0.35	0.51		0.65	1.18	
	Length-of-interview statement	0.41	0.36		0.44	0.44		0.65	0.63		1.48	1.28	

Note: Unit of analysis is calls in which an action occurred. Ns for cells are in Table 4, Panel B. See Table 3 for calculation of variables. Tests compare these means for the two call outcomes.

Cells for actions that are "incongruent" with the outcome are in gray font. Base analytic sample is matched pairs: 257 acceptances and 257 declinations. Interaction begins when sample member answers or is brought to the phone and ends when the sample member hangs up (declination) or the interview begins (acceptance).

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

study references (see summary in column (d)) and the request to participate (see summary in column (c)), although the difference is not always significant. Nevertheless, disfluency in these actions could have a cumulative effect on participation for sample members who have not exited. If we compare the two congruent actions in the recruitment phase – follow-up actions for declination calls and talk about the length of interview for acceptance calls – the latter seem to be more disfluent. However, as seen in Table 5 (column (c)), the average number of such actions that a sample member who accepts the request experiences (1.95) is fewer than the average number of follow-up actions for a sample member who declines (5.3).

The description provided in Table 6 supports our second prediction that most actions in the identification phase are more fluent than those in later phases. Perhaps surprisingly, in the recruitment phase, we do not observe that follow-up actions for declinations are more disfluent than the congruent actions for acceptances, so that our third prediction is not supported. However, our description identifies several components of the sample member's exposure to the interviewer that could be important to distinguish: Which actions occur, the number of times each action occurs, the number of words in the action, and whether the action is performed disfluently.

Table 6 suggests that when similar actions occur in declination and acceptance calls, they seem to have similar levels of disfluency, and the results of tests of differences are neither consistent nor strong. If we focus on the patterns in Table 6, we could say that for some actions – such as the request to speak to the sample member – there appear to be more disfluencies in acceptance calls; for others – notably the greeting, talk about the study, and the request to participate – disfluencies appear higher for declinations. However, only two of those differences are statistically significant. Table 6 also reinforces the observation based on Table 5 that the actions that distinguish declinations (e.g., early exits and follow-ups to declinations) and acceptances (statements about the length of the interview) make the interaction in calls that continue, in some ways, fundamentally incomparable.

6.3. Predicting Acceptance from Interviewers' Disfluencies, Words, and Actions

The structural dependencies among features of talk such as the number of actions, words, and disfluencies are reflected in correlations high enough that it is difficult to distinguish their relative contributions. For example, among the number of disfluencies, words, and actions, the correlations range from 0.70 to 0.92 considering all recruitment actions and all calls (detail not shown). In addition, all of these features can be viewed as indexing the length of the interaction, which is plausibly a product of the sample member's propensity to participate more than of the actions of the interviewer. To take the correlation between the number of words and disfluencies into account, we calculate a ratio of disfluencies per 100 words (without disfluencies); this measure is similar to the "filler rate" used by Bortfeld et al. (2001); see also Conrad et al. (2013). Table 7 presents the results of this analysis. To facilitate comparisons with previous studies, we first examine the disfluency ratio and then add the number of actions to control for the length of the interaction.

We find that when all recruitment actions are considered together, the disfluency ratio is not a significant predictor of acceptance. However, when we consider only actions up to the first turning point, there is a modest positive relationship between the disfluency ratio

Table 7. Multivariate and bivariate conditional logistic regressions of acceptance of request for participation on disfluencies, words, and actions, by location relative to first turning point.

Model	Recruitment actions	Measure	Odds ratio	p (two-tailed)	95% CI	
					Lower	Upper
Disfluency ratio only	All	Disfluencies/(words/100)	1.03	0.23	0.98	1.09
	Up to turning point	Disfluencies/(words/100)	1.06	0.03	1.00	1.11
	After turning point	Disfluencies/(words/100)	0.98	0.21	0.95	1.01
Disfluency ratio and actions	All	Disfluencies/(words/100)	1.04	0.19	0.98	1.09
		Number of actions	0.98	0.35	0.95	1.02
	Up to turning point	Disfluencies/(words/100)	1.03	0.32	0.97	1.09
		Number of actions	1.33	0.00	1.22	1.46
	After turning point	Disfluencies/(words/100)	1.00	0.96	0.97	1.04
		Number of actions	0.87	0.00	0.82	0.92
Disfluencies only	All	Number of disfluencies	1.00	0.83	0.98	1.02
	Up to turning point	Number of disfluencies	1.09	0.00	1.04	1.15
	After turning point	Number of disfluencies	0.96	0.03	0.92	1.00
Words only	All	Number of words/100	0.97	0.66	0.83	1.12
	Up to turning point	Number of words/100	4.94	0.00	2.91	8.37
	After turning point	Number of words/100	0.73	0.00	0.59	0.90
Actions only	All	Number of actions	0.99	0.43	0.95	1.02
	Up to turning point	Number of actions	1.34	0.00	1.23	1.46
	After turning point	Number of actions	0.87	0.00	0.82	0.92

Note: Unit of analysis is 257 matched pairs of acceptances and declinations. The interviewer's disfluencies, words, and actions are summed over all the recruitment actions in Table 1. For acceptances, the first turning point is the first request for participation or the last interviewer action before an acceptance. For declinations, the first turning point is the last interviewer action before the first (blocking) declination, hang up, or acceptance, whichever came first. Interaction begins when sample member answers or is brought to the phone and ends when the sample member hangs up (declination) or the interview begins (acceptance).

and acceptance ($p = 0.03$). When we add the number of actions to the models, the picture changes: The disfluency ratio no longer predicts acceptance. Instead, the number of actions before the turning point has a large positive effect on acceptance and the number after has a large negative relationship ($p = 0.00$ for both). A model that removes the structure built into a disfluency ratio by using the number of disfluencies and number of words as separate predictors shows the same result for the disfluencies (not shown).

To put these results in context, we also estimated bivariate models for each of the number of disfluencies, number of words, and number of actions (shown in Table 7). We find that for each of these three measures the relationship is null when all recruitment actions are pooled, positive for actions before the turning point, negative after. Each of these measures suggest that longer interactions before the first turning point predict acceptance, longer interactions after the first turning point predict declination.

7. Discussion

We provide a new, detailed description of how the actions of the interviewer in initial calls to recruit a sample member differ in calls that end in acceptance and declination. This description is similar in spirit to the discussion in Sturgis and Campanelli (1998), but we are able to provide more detail. This detail clarifies some of the challenges in studying how the interviewer affects participation. Our case study shows that taking the action structure of the call seriously affects conclusions: Calls that end in declinations and acceptances are most comparable only in the identification phase of the call because subsequently they consist of different actions. Levels of disfluency that occur later in the call originate in different actions, with different numbers of words, and different levels of fluency. These facts complicate the goal of predicting how the interviewer's disfluencies influence participation, because it is not clear that disfluencies that arise in different actions (e.g., scheduling interviews vs. responding to declinations) can be compared.

Our description and predictive analysis illustrate that the challenge of how to appropriately control for the different lengths and constituent actions of calls with different outcomes does not have a simple solution. This is not merely a technical issue – it potentially matters, for example, if the impact of the interviewer's talk is due to its length, its disfluency, or the actions in which the talk occurs; but these are structurally related and difficult to distinguish. We follow earlier analyses of disfluencies in considering the entire recruitment interaction in our predictive analyses (Conrad et al. 2013). However, we also compare actions up to (and including) the first turning point to those after, and control for the number of actions; we find that the turning point is important. If we predict participation, the disfluency ratio before the first turning point has a modest positive relationship with participation, but that estimate loses significance when the number of actions is included as a predictor. It is not clear whether these results reflect the results of actions by the interviewer or simply summarize a description of the call that is driven by the sample member's propensity to participate. In addition, even when we structure our analysis to consider the first turning point, we are counting disfluencies and words in different – and arguably incomparable – actions in calls with different outcomes. Thus, additional refinements to the analysis of actions are also needed to identify sites, if any, where fluency might be particularly crucial and useful to compare across outcomes.

Concepts related to, but different from, disfluencies as examined here include acoustic measures of fluency and ratings of perceived fluency (e.g., Sharf and Lehman 1984; Van der Vaart et al. 2006), following a script (e.g., Schaeffer et al. 2013), sounding scripted, or sounding "robotic" (e.g., Conrad et al. 2013). The operationalizations that accompany these various concepts include some elements not used here, such as pauses, re-starts, perceptions of listeners, and so forth. Continued work is needed to understand which of these related concepts and operationalizations, if any, enters the decisions of sample members. Understanding the impact of the fluency of the interviewer's talk is potentially important, if interviewers could be screened for or trained for fluency. The practice of giving interviewers an "agenda" rather than a script to use in recruitment (Houtkoop-Steenstra and Van den Bergh 2002; Morton-Williams 1993) may put more importance on the interviewer's ability to be fluent in a range of actions, both rote and improvised. If, for example, descriptions of the study that are more scripted and less disfluent are more effective, increasing fluency in this task could be a focus of interviewer training.

Prior studies provided a strong foundation for our operationalizations of disfluency, but we found that our interviews included complex combinations of fillers, nonpropositional elements, and other components that challenged our coding methods. It still seems possible to us that some types of disfluencies in some actions could reduce the sample member's likelihood of participating, perhaps because the disfluencies are irritating, or slow down the interviewer's delivery, or suggest incompetence. The disfluencies of the interviewer in Excerpt 2 for example, combined with his speed, certainly do not give a positive impression. However, the intuition formed by listening to such interactions has not yet led to a discovery of when and how fluency matters. We also were not able to code the great variety of complex midstream re-starts, which could be an important type of disfluency. Developing more sensitive, and potentially more informative, measures of disfluencies requires additional qualitative work.

We note that phone contacts recruiting sample members continue to be important in longitudinal and other list samples, and for a range of other purposes. By describing the actions of the interviewer in the recruitment call in more detail and differently than has been done to date (e.g., Conrad et al. 2013), we aim to deepen the way that we think about the interviewer's actions, how they depend on actions that came before, and what those dependencies imply for quantitative analyses that must summarize over such details. Our data are from a longitudinal study whose sample members are older, homogeneous in many ways, and contacted at a time when landlines were dominant, and so some details that we observe may be specific to our case. Our data were collected when landlines were still dominant, and although we believe that the trajectory of calls on cell phones differs from the trajectory we describe, we cannot say exactly how, and we are not likely to have a comparable collection of calls on cell phones anytime soon. In addition, we expect our approach to continue to be useful for analyzing actions and thinking about the challenges of determining the extent to which participation reflects the sample member's pre-existing propensity to participate versus the interviewer's action. In any case, the unusual combination of data sources (recordings, transcripts, case-control design, and participation as a criterion) provides a laboratory for exploring what such resources can teach us.

8. Appendix

Details about sample:

Schaeffer et al. (2013) give details about the sample, estimated propensity scores, and reliability of identification of actions. The 1964 data collection had an 87% response rate (http://www.ssc.wisc.edu/wlsresearch/documentation/retention/cor1004_retention.pdf). All interviews were conducted in English at whatever telephone number (usually a landline) the sample member provided. 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. In 135 of the calls in the full analytic sample of 514 cases a third party answers the telephone and calls the sample member to the telephone.

Operationalization of disfluencies:

Our operationalization of disfluencies was adapted from that of Bortfeld et al. (2001) who provide a detailed description of their method. Some of their procedures were more complex and detailed than those we had resources to implement, and so we made some adaptations and simplifications. Coding was done using string functions in Stata, supplemented by review of cases that did not match the coding rules; because we relied on machine coding, we do not estimate reliability.

Bortfeld et al. developed a complex system for identifying the location of fillers that included "phrase-internal fillers" and "between-phrase" fillers. We use a simpler system that builds on our analysis of actions to distinguish three locations for disfluencies:

1. Beginning of turn (i.e., before the first action in the turn),
2. At the beginning of a second or higher-order action in a turn. Fillers that come between actions are allocated to the later action, and
3. Within an action (including at the end of the last action in a turn).

Our counting rules were adopted or adapted from Bortfeld et al. (2001, 131–133). They code these for turns; we code for actions.

1. When one disfluency followed another (e.g., "um um" or "um uh" or "um I- ah") each was counted as a disfluency,
2. A filler was counted as beginning an action if it immediately preceded the action (whether or not the filler was itself immediately preceded by a "nonpropositional element"), and
3. Bortfeld et al. ignored "nonpropositional elements" in determining whether a filler began a turn. Our data included complex strings that combined fillers and nonpropositional elements before actions in complex ways (e.g., "and uh um"). We proceeded in these ways:
   a. In computing total initial fillers, we counted just the fillers that immediately preceded an action and were not interrupted by other elements (e.g., "um uh ACTION"),
   b. In computing total initial disfluencies, nonpropositional elements that immediately

ACTION"),

c. Broken-off talk immediately preceding an action is counted as a initial disfluency, and it is counted when computing total disfluencies (e.g., "um uh broken-off ACTION"), and

d. Midstream (mid-action) disfluencies include fillers and broken-off talk (e.g., "ACTION-begins um broken-off uh ACTION-continues")

Bortfeld et al. included the following in their word counts: "fillers, word fragments, and other words implicated in repeats and restarts." We did not include fillers or broken-off talk in our word count. We included nonpropositional elements when they were not counted as fillers. Thus, we examined two word counts: One that excluded fillers and broken-off talk, and one that excluded fillers, broken-off talk, and nonpropositional elements that were part of a string of fillers. We report analyses with the second.

When we consider all actions together, we combine actions that are first with those later in a turn.

9.   References

Arnold, J.E., C.L.H. Kam, and M.K. Tanenhaus. 2007. "If You Say Thee uh You Are Describing Something Hard: The On-Line Attribution of Disfluency During Reference Comprehension." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33(5): 914–930. DOI: https://doi.org/10.1037/0278-7393.33.5.914.

Arnold, J.E., M.K. Tanenhaus, R.J. Altmann, and M. Fagnano. 2004. "The Old and Thee, uh, New: Disfluency and Reference Resolution." *Psychological Science* 15(9): 578–582. DOI: https://doi.org/10.1111/j.0956-7976.2004.00723.x.

Barr, D.J. and M. Seyfeddinipur. 2010. "The role of fillers in listener attributions for speaker disfluency." *Language and Cognitive Processes* 25(4): 441–455. DOI: https://doi.org/10.1080/01690960903047122.

Boomer, D.S. 1965. "Hesitation and Grammatical Encoding." *Language and Speech* 8(3): 148–158. DOI: https://doi.org/10.1177%2F002383096500800302.

Bortfeld, H., S.D. Leon, J.E. Bloom, M.F. Schober, and S.E. Brennan. 2001. "Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender." *Language and Speech* 44(2): 123–149. DOI: https://doi.org/10.1177%2F00238309010 440020101.

Brennan, S.E. and M.F. Schober. 2001. "How listeners compensate for disfluencies in spontaneous speech." *Journal of Memory and Language* 44: 274–296. DOI: https://doi.org/10.1006/jmla.2000.2753.

Chamberlain, G. 1980. "Analysis of Covariance with Qualitative Data." *The Review of Economic Studies* 47(1): 225–238. DOI: https://doi.org/10.2307/2297110.

Christenfeld, N. 1995. "Does it hurt to say um?" *Journal of Nonverbal Behavior* 19(3): 171–186. DOI: https://doi.org/10.1007/BF02175503.

Clark, H.H. and J.E. Fox Tree. 2002. "Using uh and um in spontaneous speaking." *Cognition* 84(1): 73–111. DOI: https://doi.org/10.1016/S0010-0277(02)00017-3.

Conrad, F.G., J.S. 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(1): 191–210. DOI: https://doi.org/10.1111/j.1467-985X.2012.01064.x.

Dijkstra, W. and J. Smit. 2002. "Persuading Reluctant Recipients in Telephone Surveys." In *Survey Nonresponse*, edited by R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.A. Little, 121–134. New York: John Wiley & Sons.

Draisma, S. and W. Dijkstra. 2004. "Response Latency and (Para)linguistic Expression as Indicators of Response Error." In *Methods for Testing and Evaluating Survey Questionnaires*, edited by S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer, 131–148. New York: Springer-Verlag.

Draisma, S., Y. Ongena, and W. Dijkstra. 2005. "Qualified Answers and Other Doubt Expressions as Indicators of Cognitive Problems in a Health Survey." American Association for Public Opinion Research Conference, Miami Beach, FL, May 2005. Available at: http://www.asasrms.org/Proceedings/y2005f.html (accessed June 2019).

Fox Tree, J.E. 2002. "Interpreting Pauses and Ums at Turn Exchanges." *Discourse Processes* 34(1): 37–55. DOI: https://doi.org/10.1207/S15326950DP3401_2.

Fox Tree, J.E. 2007. "Folk notions of um and uh, you know, and like." *Text & Talk – An Interdisciplinary Journal of Language, Discourse Communication Studies* 23(3): 297–314. DOI: https://doi.org/10.1515/TEXT.2007.012.

Garbarski, D., N.C. Schaeffer, and J. Dykema. 2011. "Are Interactional Behaviors Exhibited When the Self-Reported Health Question Is Asked Associated with Health Status?" *Social Science Research* 40(4): 1025–1036. DOI: https://doi.org/10.1016/j.ssresearch.2011.04.002.

Groves, R.M. and M.P. Couper. 1996. "Contact-Level Influences on Cooperation in Face-to-Face Surveys." *Journal of Official Statistics* 12(1): 63–83. Available at: https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/contact-level-influences-on-cooperation-in-face-to-face-surveys.pdf (accessed June 2019).

Groves, R.M. and K. McGonagle. 2001. "A Theory-Guided Interviewer Training Protocol Regarding Survey Participation." *Journal of Official Statistics* 17(2): 249–266. Available at: https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/a-theory-guided-interviewer-training-protocol-regardingsurvey-participation.pdf (accessed June 2019).

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*, edited by J.M. Lepkowski, C. Tucker, J.M. Brick, E.D. De Leeuw, L. Japec, P.J. Lavrakas, M.W. Link, and R.L. Sangster, 385–400. New Jersey: John Wiley & Sons.

Hauser, R.M. 2005. "Survey Response in the Long Run: The Wisconsin Longitudinal Study." *Field Methods* 17(1): 3–29. DOI: https://doi.org/10.1177%2F1525822X04272452.

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,* edited by D.W. Maynard, H. Houtkoop-Steenstra, N.C. Schaeffer, and J. van der Zouwen, 205–218. New York: Wiley.

Mathiowetz, N.A. 1999. "Respondent Uncertainty as Indicator of Response Quality." *International Journal of Public Opinion Research* 11(3): 289–296. DOI: https://doi.org/10.1093/ijpor/11.3.289.

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(5): 791–814. DOI: https://doi.org/10.1177%2F0003122410379582.

Maynard, D.W. and N.C. Schaeffer. 1997. "Keeping the Gate: Declinations of the Request to Participate in a Telephone Survey Interview." *Sociological Methods & Research* 26(1): 34–79. DOI: https://doi.org/10.1177%2F0049124197026001002.

Maynard, D.W., H. Houtkoop-Steenstra, N.C. Schaffer, and J. van der Zouwen. 2002. *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*, edited by D.W. Maynard, H. Houtkoop-Steenstra, N.C. Schaeffer, and J. van der Zouwen. New York: Wiley.

Morton-Williams, J. 1993. *Interviewer Approaches*. Aldershot, England: Dartmouth Publishing.

Oksenberg, L. and C.F. Cannell. 1988. "Effects of Interviewer Vocal Characteristics on Nonresponse." In *Telephone Survey Methodology*, edited by R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg, 257–272. New York: John Wiley & Sons.

Schaeffer, N.C. forthcoming. "Interaction before and during the survey interview: Insights from conversation analysis." *International Journal of Social Research Methodology*.

Schaeffer, N.C. and J. Dykema. 2011. "Response 1 to Fowler's Chapter: Coding the Behavior of Interviewers and Respondents to Evaluate Survey Questions." In *Question Evaluation Methods: Contributing to the Science of Data Quality*, edited by J. Madans, K. Miller, A. Maitland, and G. Willis, 23–39. Hoboken, NJ: John Wiley & Sons, Inc.

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(1): 323–351. DOI: https://doi.org/10.1093/poq/nft006.

Schaeffer, N.C. and D.W. Maynard. 2002. "Occasions for Intervention: Interactional Resources for Comprehension in Standardized Survey Interviews." In *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*, edited by D.W. Maynard, H. Houtkoop-Steenstra, N.C. Schaeffer, and J. van der Zouwen, 261–280. New York: Wiley.

Schaeffer, N.C., B.H. Min, T. Purnell, D. Garbarski, and J. Dykema. 2018. "Greeting and Response: Predicting Participation from the Call Opening?" *Journal of Survey Statistics and Methodology* 1(1): 122–148. DOI: https://doi.org/10.1093/jssam/smx014.

Schegloff, E. 2010. "Some Other "Uh(m)"s." *Discourse Processes* 47(2): 130–174. DOI: https://doi.org/10.1080/01638530903223380.

Schober, M.F. and J.E. Bloom. 2004. "Discourse Cues that Respondents Have Misunderstood Survey Questions." *Discourse Processes* 38(3): 287–308. DOI: https://doi.org/10.1207/s15326950dp3803_1.

Schober, M.F., F.G. Conrad, W. Dijkstra, and Y.P. Ongena. 2012. "Disfluencies and Gaze Aversion in Unreliable Responses to Survey Questions." *Journal of Official Statistics*

28(4): 555–582. Available at: https://www.scb.se/contentassets/ca21efb41fee47d293b-bee5bf7be7fb3/disfluencies-and-gaze-aversion-in-unreliable-responses-to-survey-questions.pdf (accessed June 2019).

Sharf, D.J. and M.E. Lehman. 1984. "Relationship between the speech characteristics and effectiveness of telephone interviewers." *Journal of Phonetics* 12(3): 219–228. Abstract: https://psycnet.apa.org/record/1985-22253-001 (accessed June 2019).

Shriberg, E. 1996. "Disfluencies in Switchboard." Proceedings, International Conference on Spoken Language Processing (ICSLP '96), Vol. Addendum, 11–14. Philadelphia, PA, October 3–6, 1996. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.16.5822&rep=rep1&type=pdf (accessed June 2019).

Smith, V.L. and H.H. Clark. 1993. "On the Course of Answering Questions." *Journal of Memory and Language* 32(1): 25–38. DOI: https://doi.org/10.1006/jmla.1993.1002.

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(2): 121–139. Available at: https://search.proquest.com/docview/214805239/fulltextPDF/7DBB4A59542349A9P-Q/1?accountid=465 (accessed June 2019).

Van der Vaart, W., Y. Ongena, A. Hoogendoorn, and W. Dijkstra. 2006. "Do Interviewers' Voice Characteristics Influence Cooperation Rates in Telephone Surveys?" *International Journal of Public Opinion Research* 18(4): 488–499. DOI: https://doi.org/10.1093/ijpor/edh117.