Title
Mode and Interviewer Effects in Egocentric Network Research

Permalink
https://escholarship.org/uc/item/8fk8k6pn

Journal
Field Methods, 31(3)

ISSN
1525-822X

Authors
Fischer, Claude S
Bayham, Lindsay

Publication Date
2019-08-01

DOI
10.1177/1525822x19861321

Peer reviewed
Mode and Interviewer Effects in Egocentric Network Research

Claude S. Fischer
University of California, Berkeley
fischer1@berkeley.edu

Lindsay Bayham
University of California, Berkeley
lindsay.bayham@berkeley.edu

Acknowledgements: This research is supported by National Institute of Aging Grant R01 AG041955-01. We thank the UCNets project’s executive director, Leora Lawton, and UCNets’ many participants for support and comments; the Center for Economic and Social Research at the University of Southern California for development and data management, especially programmer Bart Orriens; and Nexant fieldwork managers, Kenric Erickson and Cat Conti.

Corresponding author: Claude S. Fischer, fischer1@berkeley.edu
Mode and Interviewer Effects in Egocentric Network Research

Abstract

Surveys of egocentric networks are especially vulnerable to methods effects. This study combines a true experiment--random assignment of respondents to receive essentially identical questions from either an in-person interviewer or an online survey--with audio recordings of the in-person interviews. We asked over 850 respondents from a general population several different name-eliciting questions. Face-to-face interviews yielded more cooperation and higher quality data but fewer names than did the web surveys. Exploring several explanations, we determine that interviewer differences account for the mode difference: Interviewers who consistently prompted respondents elicited as many alters as did the web survey and substantially more than did less active interviewers. Although both methods effects substantially influenced the volume of alters listed, they did not substantially modify associations of other variables with volume.
Mode and Interviewer Effects in Egocentric Network Research

As cost, access, and low response rates drive researchers to abandon in-person interviews for online surveys, attention has turned to understanding how mode affects respondents’ participation and answers (e.g., *Public Opinion Quarterly*, 2017). We focus here specifically on whether, how, and why online surveys and face-to-face (FTF) interviews may yield different information about egocentric networks. Such data are particularly vulnerable to procedure (e.g., Valente et al., 2017; Paik and Sanchagrin, 2013; Eagle and Proeschold-Bell, 2014; Fischer, 2012), in great measure because answering network questions is long and difficul. This study advances on earlier work by (a) randomly assigning hundreds of diverse adults to either a FTF interview or to an almost identical online survey using name-eliciting questions and (b) coding audio recordings of FTF interviews. FTF interviews yielded more respondent cooperation and higher-quality data but fewer alters than the web survey. We also found sizeable interviewer effects. Both the statistical and the recording evidence suggest that the web advantage in elicited names largely arose—despite design efforts—from the low prompting styles of some interviewers.

Previous Studies

Prior research concludes that self-administered surveys—mail or web—elicit more honest answers, while interviewers, particularly in person, yield higher response rates, induce more engagement, and more accuracy (summaries in Buelens, et al 2012, and de Leeuw and Berzelak, 2016; see also Burkill et al 2016; Kreuter et al 2008; Liu et al 2017; Gravell et al 2013; Pew 2015;
A few studies have examined mode effects specifically on network data. Matzat and Snijders (2010) found that web interviews yielded fewer names than telephone interviews. Vriens and van Ingen (2017) found no web versus FTF effects on the number named, but that web respondents reported more turnover. Bowling’s (2005) review stresses the paucity of proper experimental studies on the topic. Kolenikov and Kennedy (2014), who did use random assignment, found that web respondents provided about one fewer “close” tie than did telephone respondents. Eagle and Proeschold-Bell’s (2015) panel study of clergy found that web interviews yielded many more names (and less variance) than did telephone interviews. In sum: (a) controlled mode comparisons for network studies are rare; and (b) results are mixed but tend to show web surveys yielding fewer names than telephone surveys.

**Interviewer Effects.** Studies have found substantial interviewer effects on the number of elicited names. Paik and Sanchagrin (2013, 354) conclude that every study that has looked for them, including their own, has found interviewer effects accounting for 10 to 25 percent of the variance (e.g., van Tilburg 1998; van der Zouwen and van Tilburg 2001; Cornwell and Laumann 2013; Herz and Petermann 2017). Marsden (2003) reported interviewer effects of 15 percent in answers to the “important matters” question, “despite the extensive training of NORC interviewers and [our] high quality standards.” Moreover, Paik and Sanchagrin (2013) find that such effects explain away the controversial McPherson, et al (2006) finding of an historical loss in confidants. In sum, researchers find differences between FTF and web surveys but the direction and explanation seem uncertain. Researchers find strong interviewer effects, but how they operate is also uncertain.
The Present Study: Overview

This study adds several elements to the literature: a mode experiment with random assignment to a FTF or web condition; multiple opportunities to list alters with a variety of connections; and the audio recordings that also allow us to hear what happened in the FTF condition. We ask a sequence of research questions: First, did mode affect respondent cooperation and data quality? Second, did mode affect how many and which alters were elicited? We anticipated that web respondents would satisfice more and care about desirability less (Holbrook et al., 2003) and therefore name fewer alters. Finding, however, that online respondents gave more names led us to ask: Third, what might explain why the web condition yielded more alters? We test several explanations, including self-selection, variations in effort, and interviewer differences. Fourth, finding sizeable interviewer effects, we ask what dynamics in the FTF situation might account for them. Fifth, how much do the mode and interviewer effects influence our substantive understandings of respondents’ networks?

Data and Method

Sample. The data are drawn from the 2015 wave of the UC Nets egocentric network survey conducted in the San Francisco Bay Area (public access: https://www.icpsr.umich.edu/icpsrweb/NACDA/studies/36975/version/1). We look only at wave 1 data here because this is when the respondents were naïve. Pursuing other aims of the study, the project drew participants from two distinct age groups: 21-to-30 year-olds and 50-to-
70 year-olds, with—for this analysis—samples of 195 and 674 respondents respectively. (See Supplement, Part A.)

*Describing the Egocentric Networks.* In accord with best practices (Paik and Sanchagrin 2013), name-eliciting questions come early in the instrument. Several ask respondents to list the people with whom they are engaged in at least one of several ways: as spouse or romantic partner, household member, social companion, confidant, advisor on important decisions, practical helper, likely helper in a major emergency, recipient of the respondent’s help, and someone whom the respondent finds difficult (Supplement Part B). The 869 respondents analyzed here provided from zero (n=2) to 26 names (n=2), yielding an average of 10.2 names (IQR = 6, from 7 to 13), normally distributed. The survey then asks several questions to obtain descriptions of the alters and the ties: relationship--e.g., parent, neighbor, friend--and various descriptors--e.g., distance, gender, when met, racial homophily (Supplement Part C).

*The Mode Experiment.* During the screening interview prospective panel members were randomly assigned, at a 3:1 ratio, to either the FTF or the online condition.¹ Ninety-seven percent of the respondents who completed the study used the mode initially assigned to them. (Older women were a bit likelier than older men to end up online, but we confirmed our results by controlling for that.) The findings are the same for actual mode as intended mode, so we largely use actual mode: 647 FTF, 222 on the web.

Consistent with recommendations (Martin et al 2007), the FTF and online instruments

---

¹ Subsequent waves of the survey move some respondents from FTF to web so that we will eventually have a 1:3 ratio and will eventually be able to do both between- and within-respondent mode comparisons.
are substantively identical. The same custom-written software guided both the interviewers and the web respondents. We used a simple text questionnaire on the web to further standardize across conditions. (See further comment in the Conclusion.)

We modified the FTF condition in two general ways: When questions offered many answer options, respondents could see them on a mini-tablet screen (rather than a physical card). In addition, FTF respondents privately answered a battery of sensitive questions on the interviewers’ laptops. (The network data were complete by that point.) Critically, concerned about the possible prompting effects on web respondents of viewing several blank lines ready for names, we provided a comparable verbal prompt in the FTF condition. Interviewers were to read, “I can take up to six names” (Supplement, Part B). As we shall see, this became the leverage point for interviewer effects.

**Interviewers.** We were able to use a relatively small number of largely experienced interviewers—10 in all, five of whom completed at least 50 cases. We have audio recordings of 421 of the 647 FTF interviews. Respondents who permitted recording are not a random subset; they gave more names. Who agreed to be recorded was partly a function of the interviewer. Nonetheless, the recordings provide rare insights into the dynamics of the interviews.

**Mode Effects on Data Quality**

**Completion.** To save space, a detailed report on data quality appears as Supplement Part J. We tracked how far people who answered the invitation and began the screening process eventually got in the survey. The table in Supplement Part J presents these data, divided by age.
cohort. Young respondents were much less likely to proceed than the older respondents and they displayed a mode difference: Many fewer in the web condition continued on in each step; older respondents differed little by mode. Thus, the web interviews appear inferior in sustaining the commitment of reluctant subjects, in this case, 20-somethings (see Kreuter at al, 2008, for similar findings).

Quality. Poor quality in this context means, in particular, not adhering to instructions about naming—for example, listing couples or groups (e.g., “my family”); duplicating the same alter with a different name—for example, once as “William” and later as “Billy.” We identified that about 3-plus percent of alters listed in the FTF condition versus 13-plus percent in the web condition required hand correction.

“Pagebacks.” We can track how often respondents or interviewers paged back, presumably to correct an earlier answer, which we interpret as an effort to correct the interview. Paging back was about twice as common in the FTF mode (mn = 23 v. 13; median = 15 v. 6).

All three quality assessments suggest that the FTF condition was clearly superior.

Mode Effects on Number of Names Elicited

Network Counts. Table 1 presents the number of names—in a few configurations—that respondents gave, by mode. (A fuller table is in Supplement Part D.) There are no significant interaction effects with cohort, so we merge the data. The top line is the total number of unique alters listed across the survey (including spouses and partners). On average, web
respondents provided 1.2 more unique names, about one-third of a standard deviation more, than FTF respondents did. An alternative measure of volume is the average number of names respondents gave to the seven activity-based name-eliciting questions, from social companion through “difficult.” (The two volume measures correlate highly, $r=.78$, but differ.) The mode difference here is larger, about half an SD. In a robustness test, we trimmed and recoded outliers; the results were essentially identical.

Table 1 about here

Mode affected the number of nonkin named—1.1 more in the web condition—but not the number of kin named—only 0.2 more. One plausible explanation is that any mode effect would be weaker for core ties and that kin are more often core. However, mode affected the number of alters described as “close” ($\Delta=1.2$, $\sim\frac{1}{4}$ s.d.), so it may normative centrality rather than emotional centrality that resists a mode effect. Or, perhaps, kin are structured for more robust recall; for example, each sib remembered elicits any others (see also Brashears, 2013). Finally, the finding may reflect a primacy effect: Before the name-eliciting questions, respondents answered questions about how many close kin they had; that may have primed web and FTF respondents equally to list relatives.

The mode effect is specific to name-eliciting. We tested mode effects on 20 different items across the interview instrument (Supplement Part E). Two mode differences are significant at $p<.01$, both suggesting social desirability in the web condition. And yet: social desirability pressures should have led FTF respondents to provide more names than web respondents; they provided fewer.
The literature suggests that, despite any differences in network size, summary
descriptions of the networks should be similar across modes. We test that in Table 2 with 10
attributes of networks. The results are mixed. Six mode comparisons are not significant. On the
other hand, web respondents listed slightly higher percentages of geographically and
emotionally close alters (for reasons we could not explain). Two other measures show modest
differences: exchange multiplexity--the average number of different questions to which each
alter was named--and role multiplexity--the average number of different role labels the
respondent applied to each name. The findings imply that web respondents were more
complete in describing their networks. Later, we explore whether mode (and interviewer)
differences affected *predictors* of network attributes.

Table 2 about here

In sum, although the FTF mode elicited more respondent effort to be complete and
accurate, it elicited fewer alters—specifically, fewer *nonkin* alters—than did the web. The
mode effects are specific to name elicitation and have real but small effects on point estimates
for network attributes.

Explaining the Mode Effect

We address four explanations for the mode effect on alter volume: self-selection,
 differential effort, technical differences, interviewer effects.

*Self-Selection.* Perhaps the *dropouts* from the web condition (see above) were
disproportionately people who *would have* reported small networks had they continued.
Perhaps people uncomfortable on the web have fewer social ties. Supplement Part F describes
our efforts to test that idea by drawing using information on dropouts. The exercise provides little support for self-selection as an explanation. We also constructed a model to predict how many names the dropouts would have, based on demographics, given. This exercise, too, suggests that selective dropping out explains little, if any, of the mode effect.\(^2\)

Effort. Interviewers should be better able to encourage effort than a web program and so garner more alters (Perry et al, 2018). Although our results were the opposite, differential effort might still be a factor. A web survey allows respondents to move at their own pace and pause longer to recall names—or even, as a reviewer notes, to look them up. Perhaps, then, respondent fatigue in the FTF condition explains the fewer names. In Supplement Part G, we examine three indicators of effort: time spent on the survey; willingness to answer optional questions; and any drop-off in the number of names listed as the survey proceeded. These measures do not indicate more fatigue or less effort by the FTF respondents.

Technical Differences. One mechanical feature that might explain the mode effect is “backfilling,” respondents going back to give a name that had slipped their minds. Perhaps, FTF respondents were more reluctant to confess error and to ask the interviewer to backtrack. Paging back, as discussed earlier, was indeed associated with providing more names, but it actually occurred much more often in FTF interviews and so cannot explain the FTF deficit in names.

One technical explanation follows from literature on web survey administration: the prompting effect of seeing six blank lines on a screen may press web respondents toward giving

\(^2\) Nor are personality differences implicated. Mode is unassociated with Big-5 personality items.
more names (e.g., Vehovar et al 2008). As noted earlier, we inserted a phrase i in the FTF
version for interviewers to read each time: “I can take up to six names.” As it turns out—and as
we discuss immediately—interviewers’ following of this script was highly inconsistent, which
leads us to consider differential prompting as the major explanation for the mode effect.

Interviewer Effects

We first exam interviewer effects across our ten interviewers and afterwards ask
whether they help explain the mode effect on the volume of names. Studies suggest that
interviewer effects result from differences among interviewers in building rapport, persuading
respondents to answer sensitive questions (West and Blom 2017), probing for detail (Van der
Zouwen and Van Tilburg 2001; Houtkoop-Seenstra 1996), and getting clarifications (Mittereder
et al 2017). Other studies suggest that network differences arise if some interviewers learn to
shorten interviews by eliciting fewer names (Harling et al 2018; Josten and Trappman 2016;
Valente et al. 2017).

Our data show significant interviewer variation in the total number of unique names
elicted--from a mean of 8.2 to one of 12.2--and in the average number of names elicited per
question--2.1 to 3.7. Interviewer differences account for about eight percent of the variance in
total unique names and about 22 percent in the variance in average number of names per
question--comparable to Paik and Sanchagrin's (2013, 354) estimate from previous studies of
10 to 25 percent of the variation in total names. Adding several covariates--age, gender, race,
education, zip code--does not alter the effect sizes. And unlike some other studies, there were
no consistent time trends.

We had too few interviewers to systematically analyze what about the interviewers mattered (cf. Van Tilburg 1998; Harling et al 2017), but we found out what was actually happening in the interviews by listening to audio recordings. About two-thirds of the FTF respondents permitted us to record the interviews. Following work on differential probing behavior (West and Blom 2017; Mittereder et al., 2017; van der Zouwen and van Tilburg 2001; Houtkoop-Seenstra 1996), we examined interviewers’ probing for each of six name-eliciting questions in the survey. Did interviewers, as instructed, invite respondents to give up to six names (nine names for the sociability question)? Did they prompt respondents for names in any other way, for example, asking “Is there anyone else?” after respondents listed one or two names? We coded a random sample of 10 recordings for each of nine interviewers. Table 3 displays the average percentage of questions each interviewer accompanied with a prompt, along with other data by interviewer.

On average, interviewers read the explicit prompt (“I can take....”) only 32 percent of the time, with substantial variation between interviewers. We identified four ways that interviewer prompting influenced the number of names. First, interviewers linguistically cued how many names they expected. For example, the lowest-prompting interviewer regularly asked respondents for “a” name without telling them how many slots were available. Here are two passages.

(Respondent says that she goes to movies about once a month.)
Interviewer: "And who might you do that with?" (abbreviating the written question—see Supplement Part B).

Respondent: "My friends."

I: "Can I get a name? Or names?"

(Respondent gives two names, and interviewer moves on to the next question. The respondent may not have understood that the question concerned all sorts of social activities.)

And:

I: "Sometimes there are people in our lives who are demanding or difficult. Who are some people you find demanding or difficult?"

R: [Pause] "People’s names? Do you need a name?"

I: "Mmm-hmm."

R: "How about D.? My sister."

I: "Ok, D. it is. So, next question…"

High-prompting interviewers, in contrast, either told respondents how many slots they had available, or asked how many names a respondent wanted to list in an open-ended fashion, such as “Who do you help?”. Secondly, interviewers could prompt respondents by asking for additional names, particularly during pauses or after sets of related names (such as family). Low-prompting interviewers almost never asked for additional names. They also did not give respondents much time to list names, taking every pause in conversation as a cue to move on.
I: "Has anyone given you any practical help?... And who would that person be?"

R: [Lists a name]

I: "Very good." (Moves on to the next question.)

In contrast, high-prompting interviewers regularly asked respondents for additional names, occasionally multiple times in the same question. They also reminded respondents about specific people whom they may have forgotten. Such prompting could overcome lack of clarity about the number of slots available. For example, the interviewer who elicited the highest average number of names prompted respondents the least with how many slots remained available, but she regularly followed up by asking whether respondents had additional names.

I: “Who do you confide in?”

R: “My friend Chris.”

I: “Do you confide in K. (respondent’s husband) too?”

R: “Yeah, I guess, depending on what it is.” (laughs)

I: "Anyone else?"

R: "Not really anymore--that's about it."

And:

I: “Please think about the people that you typically do these types of things with, or other social things as well, such as..... Who are the people that you do these types of things with? I can take up to nine names.”

R: (laughs) “Well, it’s actually a much shorter list. K. L. J. (Interviewer confirms spelling after each name) And then my sister E. Oh, and then my neighbor N.
I: "Anyone else?"

R: “No, that's the main core. Oh, D.!”

I: “We got it?”

R: “Yeah.”

I: “Ok.” (*begins to move on*)

R: “Well...”

I: “I can go back.”

R: “So... ask me the question again.”

I: (*repeats the question*)

R: “So then I need to add a couple more. D. (*spells*)”

I: “I can take two more after this, and then I’m full up.”

R: “E.”

I: “I can have one more if you want.”

R: “Yeah, let me think.” (*Six-second pause*) R.

I: “Ok.”

Finally, although rare, some interviewers made explicit statements that might have discouraged respondents from giving more names.

I: “Is there anyone who’s given you practical help recently?”

R: [*pauses to think*]

I: "Probably not then, if it hasn't come to you yet."
R: "It's just hard to think. I mean we haven't moved, we haven't done anything like that."

I: "I got a 'don't know' button too."

R: "Don't know!" [Interviewer moves on]

No particular interviewer(s) can account for the mode difference. (Supplement Part H shows the results from repeated in which we dropped each interviewer’s cases.) However, prompting style—measured independently by listening to the recordings—can account for much of the mode effect. Drawing on Table 3, above, we dichotomized the FTF cases between those of interviewers who were active prompters and those of interviewers who were less active prompters. Figure 1 compares the mean number of names elicited per question for respondents of low-prompting, high-prompting interviewers, the web. (Age cohort did not interact with mode and so is not shown.) Low-prompting interviewers’ results were distinctively low, which is also true for the total number of unique names listed. The overall pattern is significant at p<.001, eta-squared = .18, but the high-prompting vs. web contrast is not significant. For the total number of unique alters, not shown, low prompting = 8.8, high-prompting = 11.2, web = 11.1 (p<.001, eta squared = .07). The high-prompting vs. web contrast is not significant. That a few interviewers could not or chose not to replicate verbally the level of prompting provided by six blank lines on a screen seems to explain the mode effect.

---

3 Combining both types of prompting in Table 4, we coded interviewers 9, 5, 8, 2, 1, and 4 (and #10) as high prompters (n = 284 respondents) and interviewers 3, 6, and 7 as low prompters (n= 363).
A reviewer asked, reasonably, whether what mattered was interviewer prompting or interviewer patience in waiting for names. The two are, of course, connected. But our eavesdropping and some limited statistical data—average time spent eliciting names is shown in Table 3—suggest that active prompting was the more important (see Supplement Part K). Also, high prompters engaged in significantly more paging back during the interviews.

We also examined the correlations with name volume of various attributes of the FTF situation—where it took place, who else was present, interruptions, and the like. A few modest associations appeared, but they neither changed the interviewer differences nor the implications for understanding the mode effect.

How Much Do Mode and Interviewer Matter?

Both mode and especially interviewer style are substantially associated with the number of names respondents provided. But how much do they affect substantive results? To wrap up our analysis, we estimated predictive models using basic demographics and then added mode and interviewer effects (low- vs. high-prompters) to the models. As shown in Supplement Part I, adding the methods variables—especially, low- versus high-prompting interviewer—increases

---

4 For example, more names per question emerged when the interview took place in a sitting room than outside a home or in a commercial venue, fewer names when there were no interruptions. Controlling for such factors does not alter the low- versus high-prompter difference. One reviewer noted that we have no comparable information about the setting in which web respondents answered. True. Perhaps a future study could ask web respondents or extract their permission to have a web camera record.
the R-squared notably. However, adding them to the models does not meaningfully change the other effects estimates nor the substantive implications—for example, that men reported about one fewer names than did women. In addition, we tested whether mode and/or interviewer affected the associations of key demographic variables with several network counts. They did not. Using two-way ANOVAs, we tested 20 interactions: the effects of method (low-prompter v. high-prompter v. web) crossed by, separately, age cohort, gender, having less than a BA degree, and partner status on the number of alters, number of nonkin alters, mean number of alters per question, number of sociability partners, and number of confidants. No interaction effects were significant at p<.05; one reached p<.10.5

Conclusion

The good news is that an online egocentric network survey can generate as many names as can a vigorous interviewer; the bad news is that less vigorous interviewers can substantially reduce the observed size of networks and that online surveys have quality issues. By combining a true experiment with listening in on interviews, we found that: (1) FTF interviews yielded more completions, cooperation, and higher quality data—although we have also found that neither mode nor interviewer affected the chances that respondents participated in the next wave (contra Pickery et al 2001). (2) FTF interviews yielded about 10 percent fewer total alters—specifically, about 15 percent fewer nonkin alters--than did the web condition and, even

5 Respondents with a spouse or partner generally reported more names (more kin, actually) than those without one, but the difference is minimal for respondents of low-prompting interviewers.
more sharply, yielded almost 20 percent fewer alters per name-eliciting question than the web (Table 2). (3) Large interviewer differences were the key to understanding the mode effect: Those interviewers who actively prompted elicited as many total names as did the blank lines on the web screen. (4) Mode and interviewer significantly affected the volume of alters elicited but did not change the observed effects of other variables on that volume or on other network counts.

As one reviewer noted, there are many new, visually interactive, online name-generator packages available (Stark and Krosnick 2017; Hogan et al. 2016; Tubaro et al. 2014; McCarty and Govindaramanujam 2005). Such formats might yield different data quality or quantity than we observed. Given that our web questionnaire was probably less engaging and motivating than such programs are, they might elicit even more names than our online format did.

The implication for practitioners is that the much cheaper web survey can yield the volume of alters that conscientious FTF interviewers can, but with trade-off of fewer completions and more errors in the data. The ideal although highly expensive option is that well-trained, -motivated, and -supervised interviewers can yield both quality data and longer lists of alters.
References


Table 1. Average number of Alters Named by Type, by Mode.

<table>
<thead>
<tr>
<th></th>
<th>Face-to-Face (n=647)</th>
<th>Web (n=222)</th>
<th>Diff.</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total N of unique alters</td>
<td>9.9 4.3</td>
<td>11.1 4.5</td>
<td>1.2</td>
<td>**</td>
</tr>
<tr>
<td>Mn. N of alters per name-elic. question (^a)</td>
<td>2.9 1.3</td>
<td>3.5 1.2</td>
<td>0.6</td>
<td>***</td>
</tr>
<tr>
<td>N listed in household</td>
<td>1.2 1.3</td>
<td>1.3 1.1</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>N kin (excl. spouse)</td>
<td>3.2 2.4</td>
<td>3.4 2.5</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>N nonkin (excl. partner)</td>
<td>6.0 3.8</td>
<td>7.1 4.2</td>
<td>1.1</td>
<td>**</td>
</tr>
</tbody>
</table>

Notes:
**Mode effects significant at p<. 01, *** p< .001 based on two-way ANOVA with cohort as a crossed factor.
\(^a\) Questions listed in Supplement Part B.
Table 2. Network Descriptions by Mode.

<table>
<thead>
<tr>
<th>Attributes of Networks</th>
<th>Face-to-Face (n=645)</th>
<th>Web (n=222)</th>
<th>Diff.</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prop. kin</td>
<td>0.38 0.25</td>
<td>0.37 0.24</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>Prop. feel close to</td>
<td>0.45 0.26</td>
<td>0.52 0.24</td>
<td>0.07</td>
<td>**</td>
</tr>
<tr>
<td>Prop. within 5 minutes</td>
<td>0.19 0.20</td>
<td>0.23 0.18</td>
<td>0.04</td>
<td>***</td>
</tr>
<tr>
<td>Prop. met in last year</td>
<td>0.08 0.16</td>
<td>0.09 0.19</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Prop. same gender</td>
<td>0.63 0.24</td>
<td>0.64 0.19</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Prop. same age</td>
<td>0.48 0.26</td>
<td>0.52 0.26</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Prop. same race</td>
<td>0.74 0.29</td>
<td>0.75 0.26</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Prop. same politics</td>
<td>0.81 0.25</td>
<td>0.83 0.22</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Mn. exch. multiplexity</td>
<td>2.09 0.65</td>
<td>2.33 0.64</td>
<td>0.24</td>
<td>***</td>
</tr>
<tr>
<td>Mn. role multiplexity</td>
<td>1.35 0.39</td>
<td>1.44 0.45</td>
<td>0.10</td>
<td>***</td>
</tr>
</tbody>
</table>

Notes:
Question texts are in Online Supplement, Section C. Mode effects significant at ** p<.01, *** p<.001. Statistical tests are based on two-way ANOVA with cohort sample as a crossed factor. There is a significant interaction effect (p<.05) for Prop. same race (.61 vs. .71 for the young; .78 vs. .77 for the older respondents).

a. The “exchange multiplexity” of each alter is the number of active name-eliciting questions (social to difficult) the alter was nominated to. The mean here is the average of those alter scores for each respondent.

b. The “role multiplexity” of each alter is the number of different kinds of relationships (e.g., sibling, coworker, friend) ego reported having with that person. The mean here is the average of those alter scores for each respondent.
Table 3: Interviewer Differences in Prompting Behavior (Based on Ten Recordings Each), Number of Names Elicited, and Mean Time in Eliciting Section.

<table>
<thead>
<tr>
<th>Interviewer</th>
<th>Average percent of questions prompteda</th>
<th>Percent of q’s using other promptingb</th>
<th>Mean number unique names elicitedc</th>
<th>Mean number of names per questiond</th>
<th>Mean number of minutes in name-eliciting sectione</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31</td>
<td>10</td>
<td>10.7</td>
<td>3.3</td>
<td>8.5</td>
</tr>
<tr>
<td>2</td>
<td>63</td>
<td>3</td>
<td>12.2</td>
<td>3.6</td>
<td>7.3</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>2</td>
<td>8.7</td>
<td>2.1</td>
<td>6.5</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>36</td>
<td>12.0</td>
<td>3.7</td>
<td>8.9</td>
</tr>
<tr>
<td>5</td>
<td>45</td>
<td>33</td>
<td>10.3</td>
<td>3.2</td>
<td>8.9</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>3</td>
<td>9.5</td>
<td>2.8</td>
<td>7.6</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>12</td>
<td>8.2</td>
<td>2.2</td>
<td>8.2</td>
</tr>
<tr>
<td>8</td>
<td>61</td>
<td>9</td>
<td>11.5</td>
<td>3.6</td>
<td>9.3</td>
</tr>
<tr>
<td>9</td>
<td>60</td>
<td>53</td>
<td>11.7</td>
<td>3.7</td>
<td>10.0</td>
</tr>
</tbody>
</table>

Notes: Interviewer 10 only conducted two interviews, so was not included in the analysis.

a. We consider interviewers “prompting” respondents when they mimicked the web administration of the survey by reading the entire prompt and—in particular—telling respondents how many names they could list for a question.

b. “Other prompting language” included instances where the interviewer asked the respondent questions such as, “Is there anyone else?”, “Is that all?”, or “How about your sister?”, or if interviewers told respondents “I have space for x more names.”

c. The average number of total unique names respondents provided to that interviewer.

d. The average, across respondents, of the mean number of names provided per question for that interviewer.

e. The average, across respondents, of the mean number of minutes in the main name-eliciting section for that interviewer. (The mean for web respondents is 9.6, with a higher skew than among interviewers.)
Figure 1. Mean Number of Names Elicited per Question Elicited for Respondents Interviewed by High-Prompting Interviewers, Interviewed by Low-Prompting Interviewers, and Surveyed on the Web.
SUPPLEMENT TO

“Mode and Interviewer Effects in Egocentric Network Research.”

Part A. More Details on Sampling and Method

Fuller descriptions of procedures, the survey instrument, the codebook and a link to the actual data are available at the project website, http://ucnets.berkeley.edu/.

Using address-based methods, we randomly sampled households from 30 census tracts randomly selected proportional to population. Solicitation letters invited any member of the sampled household who qualified by age to join the panel and be paid for participating. This recruitment procedure sufficed to generate the 50-to-70 year-old sample (n= 674 ), but attracted not enough 21-to-30 year-olds, who are notoriously hard to reach. We supplemented the young sample by asking existing respondents to invite friends, yielding 35 cases, for a total of 195 young adults randomly assigned to condition for the present analysis. (We gained another 290 completed cases through Facebook advertisements, but because all those respondents took the survey online, they are excluded from our methods analysis.) We estimate that the final panel enrollment rate, from solicitation to completion at about three percent, as comparable to the better commercial panel studies. Although for other purposes, post-stratification weights can be used to correct for an over-representation of women and the highly-educated, for the present methodological analyses, each case is equally weighted.

The fieldwork agency assigned interviewers on the basis of date and time availability, with their distance to the designated respondents as a secondary consideration. Other factors were not considered.1 In training, interviewers were permitted to deviate from the script insofar as keeping to the script would have stilted the conversation and irritated the interviewees. For example, asking respondents how long they had known their mothers seemed unwise. But this also allowed interviewers to skip text that felt repetitive in order to keep the interaction natural.

The random assignment to FTF versus web worked as planned with the older sample. Given difficulties with recruiting the younger sample, we had to adjust the assignment ratios to end up with a roughly 3:1 ratio at the end (exclusive of the Facebook recruits).

---

1 Report from the fieldwork manager (personal communication, May 18, 2017). “Once or twice, respondents requested to be assigned an interviewer of a specific gender...” [and] as many as 10 cases may have been assigned to particular interviewers “because of their skills in dealing with difficult personalities.” Moreover, “reluctance to sign up for the study or reluctance to completely answer screener questions flagged a respondent as potentially ‘difficult’ or less willing to give all responses. Also, respondents who seemed more skeptical (asked many questions regarding the legitimacy of the study or the company, etc.) were also considered to be potentially difficult cases.”
Part B. Texts of Name-Eliciting Questions (FTF version)

1. [IF MARRIED:] Please tell us your spouse’s name.

2. [IF UNMARRIED:] Is there someone whom you consider a partner or are in a romantic relationship with? [IF YES:] Please tell me this person’s name.

3. Please name all the people, besides you, who live with you at least part of the time, counting all the adults and children, if there are any.

4. Please think about people you typically do these sorts of things [leisure activities\(^2\)] with – or other social things as well, such as going shopping, out for drinks, to the park, or just hanging out. Who are the people you usually do these sorts of things with? [Include (spouse/partner name) if appropriate.] I can take up to nine names.\(^3\)

5. Sometimes personal matters come up that concern people, like issues about relationships, important things in their lives, or difficult experiences. Do you ever confide in someone about these sorts of things or do you never confide in anyone? [IF YES:] Who do you confide in about these sorts of things? [Include (spouse/partner name) if appropriate.] I can take up to six names.

6. When you have to make important decisions – for example, about taking a job, family issues, or health problems – are there any people whose advice you seek out or would seek out in making those decisions? They can be family, friends, or professional advisors. [IF YES:] Whose advice do you or would you seek out? I can take up to six names.

7. In the last few months, have any friends, relatives, or acquaintances [(who do not live with you)] given you any practical help like moving furniture, doing repairs, picking up something at the store, looking after a child, giving you a ride, or things like that? [IF YES:] Please give us the names of people who have done things like this for you in the last few months. I can take up to six names.

8. If you were seriously injured or sick and needed some help for a couple of weeks with things such as preparing meals and getting around, who would you ask?….. [IF WOULD ASK INDIVIDUALS:] Who would those people be? These can be people you have named before or new people. I can take up to six names.

9. We have been asking about people who help you out in different ways. Now, let’s turn things around. Who are the people that you help out practically, or with advice, or in other kinds of ways at least occasionally? They can be people you’ve already named or new people. I can take up to six names.

10. There are sometimes people we know who ask a lot of us, who are sometimes demanding or difficult. Who are the people that you sometimes find demanding or difficult? They can be people you’ve already named or new people. I can take up to six names.

---

\(^2\) Previous questions asked about the frequency of dining with others in homes, dining with others outside the home, and going with others to “concerts, plays, clubs, sports, or other events.”

\(^3\) These phrases, “I can take . . . names,” do not appear in the web version of the survey.
Part C. Text of Name-Descriptor Questions

As asked of all alters.

1. People can be connected to each other in a few different ways, even family members. Here is a list of the ways people can be connected. When I read a name to you, please tell me all the different ways that you are connected to that person nowadays. What are all the ways that you are connected to [NAMED OF ALTER]?
   a. Spouse/partner
   b. In a romantic relationship, but not married
   c. My parent
   d. My step-parent
   e. My child
   f. My step-child
   g. My brother/sister
   h. My step-brother/step-sister
   i. My half-brother/half-sister
   j. Other relatives, including ex’s (please specify):
   k. Housemate/roommate
   l. Neighbor
   m. Know at work
   n. Know at school
   o. Know at church, synagogue, temple, or mosque
   p. Friend
   q. Acquaintance
   r. Know another way (please specify):

2. Which of the people on this list are also [GENDER OF RESPONDENT]?

3. Which of the people on this list are about __ to __ years old? [PLUS/MINUS 6 YEARS OF RESPONDENT]

4. Which of the people on this list are about __ or older? [GREATER THAN 6 YEARS OVER RESPONDENT’S AGE]

5. Which of the people on this list did you meet just in the last year or so?

6. Which of the people on this list do you feel especially close to?

7. Which of the people on this list live in your neighborhood – say, within a 5-minute drive or so?

8. Which of the people on this list live over an hour’s drive away from you?

9. [IF RESPONDENT EVER EMPLOYED:] Which of the people on this list do/did the same kind of work as you do/did?

10. [IF R CURRENTLY IN SCHOOL:] Which of the people on this list are also going to school?

11. [IF R CURRENTLY UNEMPLOYED:] Which of the people on this list are also unemployed?

12. [IF R CURRENTLY CARING FOR HOME:] Which of the people on this list also take care at home (do not have a paid job)?

13. Which of the people on this list are of the same religion as you are?

14. Which of the people on this list are from the same racial or ethnic background as you are?

15. Which of the people on this list hold political opinions that are different from yours?
SCREENSHOTS

1. Example of name-eliciting question. This is from the FTF program. The sections highlighted in yellow do not appear in the web version.

   ![Example of name-eliciting question](image1)

2. Example of detailed name-descriptor. This is from the FTF program. The sections highlighted in yellow do not appear in the web version.

   ![Example of detailed name-descriptor](image2)
3. This is an example of a name-descriptor check-off question. This is from the FTF program. The sections highlighted in yellow do not appear in the web version. (Note that the household member, Ann, does not appear on the list.)
### Part D. Detailed Table of Number of Names Elicited by Mode

#### Average number of Alters Named by Type and Name-Eliciting Question, by Mode.

<table>
<thead>
<tr>
<th></th>
<th>Face-to-Face (n=647)</th>
<th>Web (n=222)</th>
<th>Diff. Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>SD</td>
<td>mean</td>
</tr>
<tr>
<td><strong>Global Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total N of unique alters</td>
<td>9.9</td>
<td>4.3</td>
<td>11.1</td>
</tr>
<tr>
<td>Mn. N of alters per name-elic. question*</td>
<td>2.9</td>
<td>1.3</td>
<td>3.5</td>
</tr>
<tr>
<td>N listed in household</td>
<td>1.2</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>N kin (excl. spouse)</td>
<td>3.2</td>
<td>2.4</td>
<td>3.4</td>
</tr>
<tr>
<td>N nonkin (excl. partner)</td>
<td>6.0</td>
<td>3.8</td>
<td>7.1</td>
</tr>
<tr>
<td><strong>Counts for Specific Questions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N named as social companion</td>
<td>5.4</td>
<td>2.8</td>
<td>6.2</td>
</tr>
<tr>
<td>N named as confidant</td>
<td>3.0</td>
<td>1.9</td>
<td>3.6</td>
</tr>
<tr>
<td>N named as advisor</td>
<td>2.5</td>
<td>1.8</td>
<td>3.4</td>
</tr>
<tr>
<td>N named as practical helper</td>
<td>1.7</td>
<td>1.7</td>
<td>2.2</td>
</tr>
<tr>
<td>N named as emergency helper</td>
<td>2.8</td>
<td>1.7</td>
<td>3.5</td>
</tr>
<tr>
<td>N named as recipient of help from respond.</td>
<td>3.7</td>
<td>1.8</td>
<td>4.5</td>
</tr>
<tr>
<td>N named as difficult</td>
<td>1.2</td>
<td>1.2</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Notes: Mode effects significant at ** p<.01, *** p<.001. Statistical tests are based on two-way ANOVA with cohort sample as a crossed factor.

* Only includes questions listed in table, not questions 1-3 in Part B above.
Part E. The Mode Effect is Specific to Network Results.

We tested mode effects on 20 different items across the interview instrument. Two mode differences are significant at p<.01 (compared to, for example, eight for cohort differences). These two—dining with others and wishing to know more people—fit the expectation that online respondents are likelier to confess to socially undesirable traits, but other social desirability answers did not show mode effects. Perhaps any desirability effects in this survey revolve specifically around personal ties, given its purpose. (The invitation to join the study read: “This research is about our social lives . . . at how changes in social connections happen and how they affect health. . . .”) And yet: social desirability pressures should have led FTF respondents to provide more names than web respondents; but they provided fewer.

<table>
<thead>
<tr>
<th>Means or Proportions for Twenty Assorted Questions, by Mode.</th>
<th>Face-to-Face (n=647)</th>
<th>Web (n=222)</th>
<th>Diff.</th>
<th>Test²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prop. Anyone move in with respondent recently</td>
<td>.15 .36</td>
<td>.22 .41</td>
<td>.07 *</td>
<td></td>
</tr>
<tr>
<td>Prop. Have a pet</td>
<td>.48 .50</td>
<td>.48 .50</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Mn. Freq. dine with others in home (high to low)b</td>
<td>2.48 1.30</td>
<td>2.77 1.26</td>
<td>.29 **</td>
<td></td>
</tr>
<tr>
<td>Prop. Confident that family would help in emergency</td>
<td>.72 .45</td>
<td>.72 .45</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Prop. Who wish they knew more people who helpc</td>
<td>.23 .42</td>
<td>.35 .48</td>
<td>.12 ***</td>
<td></td>
</tr>
<tr>
<td>Prop. Have anything more to say about networks</td>
<td>.35 .48</td>
<td>.32 .47</td>
<td>-.03</td>
<td></td>
</tr>
<tr>
<td>Prop. Report any trouble paying bills</td>
<td>.17 .38</td>
<td>.20 .40</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Prop. Report any recent good economic event</td>
<td>.51 .50</td>
<td>.46 .50</td>
<td>-.05</td>
<td></td>
</tr>
<tr>
<td>Prop. Anyone close to respondent die in last year</td>
<td>.37 .48</td>
<td>.35 .48</td>
<td>-.02</td>
<td></td>
</tr>
<tr>
<td>Prop. Experienced a break in relationship in last year</td>
<td>.37 .43</td>
<td>.33 .42</td>
<td>-.04</td>
<td></td>
</tr>
<tr>
<td>Mn. Self-rated health (excellent to poor)</td>
<td>2.36 1.07</td>
<td>2.36 1.04</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Prop. Reported no serious health issues</td>
<td>.69 .46</td>
<td>.65 .48</td>
<td>-.04</td>
<td></td>
</tr>
<tr>
<td>Mn. Nmbr. nights per week trouble falling asleep</td>
<td>1.34 2.27</td>
<td>1.41 1.94</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>SAQ: Respondent’s self-reported weight (lbs.)</td>
<td>166.85 43.18</td>
<td>163.31 38.35</td>
<td>-3.54</td>
<td></td>
</tr>
<tr>
<td>SAQ: Mn. Nmbr. days a week respondent felt happy</td>
<td>4.52 2.30</td>
<td>4.30 2.36</td>
<td>-.22</td>
<td></td>
</tr>
<tr>
<td>SAQ: Prop. Have had thoughts of suicide</td>
<td>.05 .21</td>
<td>.04 .20</td>
<td>-.01</td>
<td></td>
</tr>
<tr>
<td>Prop. Volunteer &quot;other&quot; as their racial identity</td>
<td>.09 .29</td>
<td>.05 .23</td>
<td>-.04</td>
<td></td>
</tr>
<tr>
<td>Prop. Say they are members of religious organization</td>
<td>.31 .46</td>
<td>.28 .45</td>
<td>-.03</td>
<td></td>
</tr>
<tr>
<td>Mn. Attendance at religious services (high to low)</td>
<td>3.61 1.49</td>
<td>3.73 1.53</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>Prop. Willing to answer short surveys later</td>
<td>.96 .21</td>
<td>.96 .20</td>
<td>.00</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The questions are listed here in their order in the survey instrument.

a. ANOVA for continuous or ordinal variables; Chi-Square for dichotomies. * p<.05, **p<.01, *** p<.001.

b. There were, however, no significant differences in the frequency of dining with others in public nor in going to public events like concerts with others.

c. Web respondents were also likelier to say that they wished they knew more people to see socially 59 percent versus 41 percent, p<.001.

d. “SAQ” refers to a self-administered question. In the FTF condition, interviewers handed the laptop to respondents to answer a set of more private questions.
Part F. Assessing Differential Self-Selection as a Cause of Mode Differences

This analysis estimates to what extent differential self-selection explains why more alters were named by web than by FTF respondents. Such self-selection would entail dropouts from the web condition having fewer actual alters in their lives than those assignees to the web respondents who did complete the survey. To address the issue, we used two strategies.

The first one involved isolating web-assigned respondents who dropped out and simulating what their network size (measured as mean number of names provided per name-eliciting question) would have been had they completed. We tried three estimating procedures. In one, we assumed that these people would have been like the roughly dozen who had completed the survey but had insisted on being transferred from the web to the FTF condition. In the second, we assumed that the dropouts from the web condition were like the completers who had told the screening interviewer either that they had no access to the internet or that they were uncomfortable with computers. In the third, we assumed that the web dropouts were like the respondents who did not say, in later part of the survey, that they used email to contact their families. The results of these exercises (available from the authors) based on the idea that self-selection by technology aversion would have reduced the difference in mean number of names by no more than 14 percent, leaving a substantial mode gap to be explained. The main reason that the simulations mattered little is that so few of the older respondents assigned to the web condition failed to finish the name-eliciting section on the web (n=28).

We also considered the possibility that the personalities of the dropouts interacted with mode. Perhaps shyer, less social respondents in the end preferred not to do the survey, but those assigned to a personal interviewer found it harder to opt out than those on the web, so that the remaining web respondents were unrepresentatively high in network size. But the average personality profile of the respondents did not differ by mode. 4

Finally, we tested the possibility that the dropouts from the web mode were disproportionately people with small networks by estimating their expected number of names using our screener survey. The survey vendor asked a set of questions to would-be respondents who answered the call for participation and used the answers to determine eligibility and mode assignment. For this analysis, 227 qualified would-be respondents received a mode assignment (174 to FTF and 53 to the web) but did not end up completing the main survey and thus do not appear in the final sample. 5 We know several things about these dropouts that we also know about the respondents who completed the study, 641 of those assigned to FTF and 211 to web. 6 We know age, gender, household structure, access to the internet, and zip code.

Using these shared variables, we ran an OLS regression to “predict” the mean number of names per name-eliciting question for the 852 completed cases. The resulting equation explained about 10 percent of the variance. 7 Using the model parameters, we estimated the expected mean number of names per question for the incomplete cases. We compared the estimates for web dropouts to estimates for FTF dropouts. Had the web-assigned dropouts differed from the FTF-dropouts in gender, age, and other ways associated with listing many names, the two estimates would have differed. They did not: The

---

4 The survey employed a short version of the “Big 5” personality test.
5 As in all analyses, we exclude the respondents recruited through Facebook.
6 These numbers exclude a few cases with missing data.
7 The predictors were age*, age-squared**, whether there was 21-30 year-old in the household, whether there was a 50-to-70 year-old in the household, the number of household members in the (would-be) respondent’s age bracket, being female***, internet access***, and eight dummy variables for zip code truncated to the first three digits** (where the asterisks indicate significance).
projected mean number names per question was 3.04 for the would-be web respondents and 3.01 for the would-be FTF respondents. (By the way, there was also no difference in the projections among the respondents who in the end completed the study—3.04 v. 3.0—another way of showing that the mode effect was not spurious, at least not spurious because of the selection associated with variables included in this model.)
Part G. Assessing Differential Effort as a Cause of Mode Differences

We assess the possibility that differences in effort between FTF and web respondents explain the differences in the number of alters they listed by examining three indicators of effort: time spent on the survey, willingness to answer optional questions, and a decline in the number of names listed as the survey proceeds.

*Duration.* The total time the survey took can reflect effort—how much respondents were willing to invest in the project. To be sure, duration reflects other factors as well, such as respondents’ comprehension, distraction, and verbosity. But time serves as one gauge of effort. The results are shown in the table below, top line. On average, web respondents spent 4.3 minutes more on the survey than FTF respondents did (p<.05). However, the variance was much higher for web respondents; they included more high-side outliers. Indeed, FTF cases could not have extreme outliers because interviewers were instructed to just stop after two hours. When we control for outliers by trimming the 5th and 95th percentiles, the new results (line 2), show only a 1.2 minute difference. We conclude that (a) the web condition allowed for extremely long interviews, which the FTF condition did not; but (b) in general, there was no substantial difference in time spent between the two modes.

*Answering Optional Questions.* Another indicator of effort or commitment is whether a respondent is willing to answer optional questions, particularly ones that might open the door to yet more probing. We used three such questions: asking respondents an open-ended question about whether any other economic event had occurred to them in the past year (beyond those we had explicitly asked about); asking similarly about another life event that we had not specifically stipulated; and asking whether they had any health limitation that we had not specifically asked about. Answering yes to any of these required an additional answer and an explanation. The third to fifth lines in the table show, effectively, no difference by mode. Respondents who answered yes to these questions tended also to give more names, perhaps because both are indications of respondent effort, but that was unrelated to mode, suggesting that the mode effect on alter volume is not explained a mode effect on effort.

### Measures of Respondent Effort by Mode.

<table>
<thead>
<tr>
<th></th>
<th>Face-to-Face (n=647)</th>
<th>Web (n=222)</th>
<th>Diff.</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mn. time of survey (minutes)</td>
<td>60.3 / 15.6</td>
<td>64.7 / 26.7</td>
<td>4.3</td>
<td>*</td>
</tr>
<tr>
<td>Mn. time of survey, trimmed(^a)</td>
<td>60.2 / 14.3</td>
<td>61.9 / 17.6</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Pct. offered other economic life event(^b)</td>
<td>36.4 / 14.3</td>
<td>42.3 / 17.6</td>
<td>6.0</td>
<td></td>
</tr>
<tr>
<td>Pct. offered any other life event(^c)</td>
<td>34.6 / 14.3</td>
<td>33.1 / 17.6</td>
<td>-1.5</td>
<td></td>
</tr>
<tr>
<td>Pct. offered additional health limitation(^d)</td>
<td>25.2 / 14.3</td>
<td>28.8 / 17.6</td>
<td>3.6</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Mode effects significant at *p<.05. Statistical tests for means are based on two-way ANOVA with cohort sample as a crossed factor. Statistical tests for percentages are Fisher exact tests, also tested with cohort as a control.

a. “Trimmed” by re-setting all values under the 5th percentile (of both treatments pooled) to the value of the 5th percentile and similarly with values above the 95th percentile.

b. “In roughly the last year or so, have you had another important work or economic event happen to you, for good or bad?”
c. “Are there any other important events, both good and bad, that have happened to you in the
last year or so that we have not covered?” Note that in the 21-to-30 year-old sample, there is a
significant ($p<.05$) difference: 33.1% in FTF and 17.9%.
d. “Is there any other aspect of your health that limits your activity in some way that we haven’t
yet discussed?”

*Waning of Effort.* Perhaps the FTF-web difference emerged as all respondents realized how
taxing the procedure is and the FTF respondents—assuming that their experience was more stressful and
fatiguing—began curtailing names earlier than the web respondents did, thus producing the observed
differences. We return to Part D in this supplement. The specific questions are listed in their order within
the interview. One observes a drop in the mean number of names provided from the first to the fourth
question, but then an increase back to about the initial level by the sixth question, closing finally with the
lowest mean. More important, however, the FTF-web difference in names provided did not change
systematically as the survey proceeded.

In our analysis of completion rates, we looked at differential dropout after the name-eliciting
questions (see table in Section J). We can ask what percent of those respondents who answered the
name-eliciting questions did not then complete the remainder of survey—perhaps because they were too
worn out by the procedure—and whether there is a mode effect there. Among the young respondents, the
proportion dropping out after the name-eliciting section was greater for web than for FTF respondents, 13
versus 2 percent. The gap was not as wide, 7 versus 4 percent, among the older respondents. However, if
these differences in dropout rates indicate differential fatigue, they suggest that the web respondents
were more fatigued by the name-eliciting section than were the FTF respondents. And yet they gave more
names, contradicting a fatigue explanation of the mode effect. Giving many names may well have been
fatiguing (and thus encouraging of stopping), but that mechanism would not explain why web
respondents gave more names to start with.

Simulated mean names per question elicited with different interviewer compositions (FTF only; n= 647).

<table>
<thead>
<tr>
<th>Interviewer Cases Replaced</th>
<th>Mean # names per quest. in simulated FTF sample</th>
<th>Web (m=3.5) minus simulated FTF Difference</th>
<th>% change in difference from full sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without I. # 1</td>
<td>2.8</td>
<td>0.7</td>
<td>08%</td>
</tr>
<tr>
<td>Without I. # 2</td>
<td>2.8</td>
<td>0.7</td>
<td>04</td>
</tr>
<tr>
<td>Without I. # 3</td>
<td>3.0</td>
<td>0.5</td>
<td>-17</td>
</tr>
<tr>
<td>Without I. # 4</td>
<td>2.8</td>
<td>0.7</td>
<td>05</td>
</tr>
<tr>
<td>Without I. # 5</td>
<td>2.9</td>
<td>0.6</td>
<td>03</td>
</tr>
<tr>
<td>Without I. # 6</td>
<td>2.9</td>
<td>0.6</td>
<td>-01</td>
</tr>
<tr>
<td>Without I. # 7</td>
<td>3.1</td>
<td>0.4</td>
<td>-30</td>
</tr>
<tr>
<td>Without I. # 8</td>
<td>2.8</td>
<td>0.7</td>
<td>17</td>
</tr>
<tr>
<td>Without I. # 9</td>
<td>2.8</td>
<td>0.7</td>
<td>08</td>
</tr>
<tr>
<td>Without I. # 10</td>
<td>2.9</td>
<td>0.6</td>
<td>00</td>
</tr>
<tr>
<td>Without I. # 3 &amp; #7</td>
<td>3.3</td>
<td>0.2</td>
<td>-59</td>
</tr>
<tr>
<td>All Interviewers (none dropped)</td>
<td>2.87</td>
<td>0.66</td>
<td></td>
</tr>
</tbody>
</table>

* p<.05, ** p<.01, ***p<.001
Part I. Testing Implications of Mode/Interviewer Effects on Results.

Regressions of Total Number of Names, Number of Nonkin Names, and Mean Names per Question with and without Mode/Interviewer Style (n=868 unweighted respondents).

<table>
<thead>
<tr>
<th></th>
<th>Total N of Names</th>
<th>N of Nonkin</th>
<th>Mn. Names per Q.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(1)</td>
</tr>
<tr>
<td>Constant</td>
<td>7.80***</td>
<td>8.67***</td>
<td>6.60***</td>
</tr>
<tr>
<td>R is Male</td>
<td>-1.04***</td>
<td>-0.97***</td>
<td>-0.78**</td>
</tr>
<tr>
<td>Less than B.A. Degree</td>
<td>-1.72***</td>
<td>-1.80***</td>
<td>-1.78***</td>
</tr>
<tr>
<td>Older Age Cohort</td>
<td>0.87</td>
<td>0.75</td>
<td>0.53</td>
</tr>
<tr>
<td>R has Spouse or Rom. Partner</td>
<td>0.15</td>
<td>0.23</td>
<td>-0.62*</td>
</tr>
<tr>
<td>Mother Alive</td>
<td>0.41</td>
<td>0.31</td>
<td>0.07</td>
</tr>
<tr>
<td>Father Alive</td>
<td>0.93*</td>
<td>0.92*</td>
<td>0.54</td>
</tr>
<tr>
<td>N of Adult Children</td>
<td>0.33**</td>
<td>0.33**</td>
<td>-0.22</td>
</tr>
<tr>
<td>R Employed Full Time</td>
<td>0.54</td>
<td>0.46</td>
<td>0.42</td>
</tr>
<tr>
<td>N of Others in Household</td>
<td>0.92***</td>
<td>0.90***</td>
<td>0.20</td>
</tr>
<tr>
<td>R is Black</td>
<td>-0.89</td>
<td>-0.62</td>
<td>-1.16*</td>
</tr>
<tr>
<td>R Has Chronic Health Issue</td>
<td>0.64*</td>
<td>0.58*</td>
<td>0.20</td>
</tr>
<tr>
<td>Method: Low-Prompting Int.</td>
<td></td>
<td>-1.95***</td>
<td>-1.46***</td>
</tr>
<tr>
<td>Method: High-Prompting Int.</td>
<td></td>
<td>0.39</td>
<td>0.02</td>
</tr>
<tr>
<td>Method: Web (Contrast)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in R-squared</td>
<td>.06***</td>
<td>.03***</td>
<td>.16***</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>.14***</td>
<td>.20***</td>
<td>.11***</td>
</tr>
</tbody>
</table>

Notes:

1. N of Nonkin excludes romantic partners (who are counted in the independent variable, “R. has Spouse or Rom. Partner”).
2. Method: FTF Respondents are coded by how much prompting their interviewers did and contrasted against respondents on the web.
3. In regression for total names, N in household has almost automatic effect because all co-residents are listed among the network names.
4. OLS regression is used because the dependent variables are roughly distributed normally.

Significance: * p<.05, ** p<.01, *** p<.001.
Part J. Assessing Data Completion and Quality

Completion. Interviewers can presumably coax respondents into continuing a survey to the end; our data are partially consistent. We start with the people who answered the invitation to be part of the panel, began the screening process, and got as far as being randomly assigned to mode. We then track how far they got. The table below presents these data, divided by age cohort, because cohort strongly conditioned this process.

Young respondents were much less likely to continue on than were older respondents and young respondents displayed a mode difference: Many fewer of those assigned to the web condition continued on than of those assigned to the FTF condition, at each step in the process. Continuation rates differed little by mode assignment among the older respondents. Web interviews may be inferior in sustaining the commitment of reluctant—in this case, 20‐something—subjects (see Kreuter at al, 2008, for similar findings).

Quality. There should be fewer mistakes in the FTF condition—”mistakes” in this context meaning, in particular, not adhering to instructions that each response to a name‐eliciting question should be the names or initials of single individuals listed separately. Common errors were to list couples or groups, such as “Bill and Alice” or “my family,” or to duplicate the same alter with a different name, such as listing “Bill” in a different question as “William.” We identified that about 3 percent of names offered face‐to‐face versus about 13 percent of names listed on the web required correction or deletion. The audio recordings reveal interviewers indeed clarifying and correcting the listed names.

“Pagebacks.” We are able to track for both FTF and web surveys the extent to which respondents or interviewers paged back during the survey or interview, presumably to correct an earlier answer. We interpret the number of page‐backs as a measure of the effort to correct, perhaps make more comprehensive, the interview. Paging back was about twice as common in the FTF than online modes (means of 23 and 13 times; medians of 15 and 6 times).(As noted in the text, high‐prompting interviewers engaged in more page backs than low‐prompting ones [mns = 27.4 vs. 19.5, p<.01; difference in mean of logged pagebacks, p<.001].)

In all three of these quality assessments—completion, errors, and pageback efforts to be accurate—the FTF condition was clearly superior to the online condition.

[table on next page]
Completion of the Survey by *Intended* (Assigned) Mode, by Age Cohort.

<table>
<thead>
<tr>
<th></th>
<th>21-to-30 Year-Olds</th>
<th>50-to-70 Year-Olds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress of Respondent from Initial Contact with the Research</td>
<td>Face-to-Face</td>
<td>Web</td>
</tr>
<tr>
<td>N Who reached mode assignment</td>
<td>224</td>
<td>151</td>
</tr>
<tr>
<td>Pct. Who started the survey</td>
<td>65.2</td>
<td>43.0***</td>
</tr>
<tr>
<td>Pct. Answered name-eliciting quest’s(^a)</td>
<td>64.7</td>
<td>39.7*</td>
</tr>
<tr>
<td>Pct. Completed entire survey(^b)</td>
<td>63.8</td>
<td>34.4***</td>
</tr>
<tr>
<td>Final N</td>
<td>143</td>
<td>52</td>
</tr>
</tbody>
</table>

Notes:

\(^a\) Before this section were about two dozen questions largely about marital status, family composition, employment, and residential mobility.

\(^b\) The remaining sections asked name-descriptor questions, questions asking for evaluations of the networks, and then batteries largely about health, life events, and sociodemographic background.

The difference between modes in self-selection in making this step is significant, by Fisher exact test, at: *** p<.001 or * p<.05
Part K. Prompting or Patience?: Between-Interviewer Differences.

In response to a reviewer’s reasonable question that perhaps it was interviewer variation in *patience* to wait for respondents to offer up names rather than in prompting behavior that mattered, we took a closer look.

First, our direct observations from listening to 100 audio recordings is that prompting and patience tended to go together. However, between the two, we clearly observed more variation in probing and more consequential interviewee behavior from probes than from silences.

Second, we conducted a rough statistical analysis drawing from the data in Table 3, using the minutes spent in the main name-eliciting section as a rough index of patience. We looked at the associations across the nine interviewers (those with sufficient numbers of interviews) among the averages in the table. The best predictor of the average number of alters per question (and overall total) elicited across interviewers was their average rate of compliance with reading the written prompts, not by the time they spent in the name-eliciting section. (Time was best predicted by how many spontaneous, other sorts of prompts they offered.) One example of this pattern is that the least-probing interviewer, the one who also elicited the fewest names, took an average amount of time during the name-eliciting section. There was time-taking chatter going on, just not name-eliciting chatter.