Age Matters: Sampling Strategies for Studying Digital Media Effects*

Kevin Munger, Mario Luca, Jonathan Nagler, Joshua Tucker

December 21, 2018

Abstract

There is increasing evidence for age-related heterogeneities in how people use the internet and social media for politics. Convenience samples that do not ensure sufficient variation in this crucial covariate cannot be used to generalize results to any relevant population. In particular, Amazon's Mechanical Turk is the most prominent source of subjects for these experiments, but a vanishingly small number of these subjects are over 65 years old. The problem is worse for "digital literacy" (strongly correlated with age in practice but which is a theoretically distinct treatment moderator): 100% of subjects recruited via Mechanical Turk are above a threshold of digital literacy below which there are many internet users. We argue for the use of Facebook advertisements to recruit subjects which vary along this dimension, but caution that research with older adults can pose novel ethical concerns.

*This research is supported by the John S. and James L. Knight Foundation through a grant to Stanford University's Project on Democracy and the Internet.

1 Age

A slew of recent research on digital media effects presents a striking empirical regularity: older people use and experience the internet differently, in ways that have important political consequences.

The difference between "digital natives" and "digital immigrants" has largely focused on how (and to what extent) young people are distinct in their fluency with digital technology (Boyd, 2014; Prensky, 2001). This may have largely been a function of who actually used these technologies; very few older people used social media before around 2008.

Today, however, the fastest growing population of Facebook users is adults over 65 years old (Smith and Anderson, 2018). And their presence is increasingly showing up in empirical analyses.

Using web-tracking software, Guess, Nagler, and Tucker (2018) report that "Users over 65 shared nearly 7 times as many articles from fake news domains as the youngest age group." during the 2016 US Presidential election.

Analogous results from Twitter are presented by Barbera (2018), who finds that people over 65 shared roughly 4.5 as many false news stories on Twitter as people 18 to 24 (result read from a coefficient plot).

Even the most famous digital voting experiment, conducted in concert with Facebook in 2010, finds similar results (the heterogeneous effects from the experiment are reported in (Bond et al., 2017)): "The [Facebook GOTV experiment] effect size for those 50 years of age and older versus that of those ages 18 to 24 is nearly 4 times as large for self-reported voting and nearly 8 times as large for information seeking."

These are not small heterogeneities, and these are not underpowered studies grasping for statistical significance. Heterogenities of this size have serious implications for the study of online behavior, suggesting that average treatment effects may be misleading and that novel research designs that take care to understand the experiences of older internet users.[1]

Thankfully, literatures in fields complementary to Political Science offer helpful insights. Sociologists and Communications scholars (and Computer Scientists in the

[1]There is mounting evidence that many types of research typically conducted on using digital samples are low in effect heterogeneity and thus can be reasonably generalized to the population of interest (Coppock, 2018; Coppock, Leeper, and Mullinix, 2018; Snowberg and Yariv, 2018), but this evidence does not imply that effect heterogeneity does not exist. Demonstrates of generalizability do not generalize. The cited papers are careful to make this point, but it bears repeating.

subfield of Human-Computer Interaction) have theory and evidence pertaining to user experience, rather than simply analyzing trace data or survey responses. The theory of "digital literacy" is particularly helpful—it is highly (negatively) correlated with age, but it remains to be seen how much of the results above are due to digital literacy or age per se.[2]

This note aims to apply these insights to narrow but important area of political science research: the use of digital convenience samples (most famously, Amazon's Mechanical Turk) to generate results that researchers claim to be representative of some larger population. "MTurkers" are a useful sample for many forms of experiments—those in which effect heterogeneity among the target population is low. However, a vanishingly small number of MTurkers are over 65 years old. The problem is worse for "digital literacy": 100% of subjects recruited via Mechanical Turk are above a threshold of digital literacy below which there are many internet users.

We diagnose this issue in the context of an experiment aiming to measure the effect of clickbait headlines. Our solution was to recruit an alternative set of subjects using Facebook ads; these subjects tended to be much older and less digitally literate. However, many of them were unable to complete the survey, and some expressed frustration with the process. Online recruitment of low digital literacy subjects poses novel ethical challenges which we discuss in the conclusion.

2 The generalizability of online survey experiments

There have been hundreds of experimental studies conducted using subjects recruited via Amazon's Mechanical Turk (MTurk). These studies are valid insofar as treatment effects estimated on this population generalize to a population of interest; although the subject pool is not representative of the US population, the way they respond to experimental stimuli is informative when covariate reweighting is employed.

Mullinix et al. (2015) provides theoretical and empirical justification for this practice, but the authors are careful to maintain the continued importance of nationally representative samples, particularly when a given treatment has heterogeneous effects. If researchers have insufficiently theorized the dimensions of effect heterogeneity, data

[2]The source (and thus durability) of the extant digital literacy gradient is a crucial question for future research. If we are in a period of disequilibrium, and digital technology is a once-in-a-century invention, then we should expect the gradient to decrease. However, if technological progress is permanently accelerated, older people will always face novel technology, and the gap may worsen.

from nationally representative samples can help reveal them : "If one has a well-developed theory about heterogeneous treatment effects, then convenience samples only become problematic when there is a lack of variance on the predicted moderator...[eg] MTurk when a moderator is religion (i.e. MTurk samples tend to be substantially less religious than the general population)" (Mullinix et al, p123).

As the internet and social media become increasingly integrated into politics, more scholarly attention has been dedicated to the study of online political behaviors: how do people engage with politics online? Of particular interest are online behaviors which have no offline analogues.

Subjects recruited from MTurk may be inappropriate to study these behaviors. The sociologist Eszter Hargittai has advanced the theory of *digital inequality* to argue that—even among individuals who use the internet frequently and persistently—inequality in their levels of online skills ("digital literacy") has important implications for *how* they use the internet (DiMaggio, Hargittai et al., 2001; Hargittai, 2001).

This issue has become increasingly relevant as the population of internet and social media users has expanded beyond tech-savvy early adopters to encompass the majority of the US population. The fastest growing population of Facebook users are adults over 65 years old (Smith and Anderson, 2018); these individuals also tend to have much lower digital literacy (Hargittai, Piper, and Morris, 2018).[3] The study of digital literacy with survey instruments is a difficult task because the underlying technology is rapidly changing and needs to be validated against behavioral data.

In the current paper, we operationalize digital literacy through several types of "forensic" data from the online survey, including asymmetric attrition from the sample, time spent on confusing questions, and responses to an free response question (we defend these measures below). We also have direct data on a related and far less ambiguous covariate: age.

The evidence above demonstrates that online behaviors are strongly heterogenous in age. As we demonstrate below, age does not sufficiently vary within the MTurk population, making this population inappropriate to study online treatment effects for which age is a moderator. But the problem is worse for digital literacy. The site is sufficiently difficult to use that there is a structural barrier to recruiting *any* subjects

[3]Even more dramatic has been the experience of Facebook use by people in developing countries for whom Facebook is their first and only means of using the internet. In several Asian nations, the problem of racial violence inspired by false information spread via Facebook has become widespread Beech and Nang (2018).

below a certain threshold.

3 MTurk Requires Digital Literacy

The age skew of the MTurk population is a problem (Huff and Tingley, 2015), but one that may be solvable. It is possible to selectively recruit older MTurkers or reweight the data.[4] However, as Mullinix et al. (2015) argues, reweighting fails when the joint distribution properties of a sample do not match the population: "[MTurk] may have similar percentages of older individuals and racial minorities, but may not match the population based sample with respect to older minorities" (p123). This is also the case with age and digital literacy: the older people on MTurk are more digitally literate than the older people not on MTurk, meaning that there is *zero support* in this population for exactly the demographic of people most likely to have shared Fake News during the 2016 election (Barbera, 2018; Guess, Nagler, and Tucker, 2018).

Our confidence in this claim comes from Brewer, Morris, and Piper (2016), who conduct a survey of older Americans and discover that the largest barrier to participation in MTurk is that they are unaware of it. Their "survey data confirm that even among online older adults, those who have tried crowd work are (relatively) younger and more tech savvy than those who have not" (p8).

Brewer, Morris, and Piper (2016) recruit a small sample of older adults who have not used MTurk and encourage them to perform example tasks on the platform.

The vast majority of this sample of adults over 65 reported having used the internet for more than 15 years and being comfortable using computers (p2250). However, the MTurk interface proved an insurmountable:

Many participants were not familiar or comfortable with opening content in new tabs/windows, resulting in questions such as, "How do I get back to the instructions?" (P7) after a new tab was opened. Also, participants often forgot the instructions immediately upon opening the new window, particularly long and detailed instructions. (p2251)

Because of the reputation system around which MTurk operates, even if low digital literacy individuals sign up, they are likely to be excluded from future samples that require MTurkers to have maintained a certain rating.

[4]There are basically 0 people over 75 (the "oldest old") on the platform, but it is unknown how excactly age functions *within* the set of older people.

Qualitative study of the way that different populations engage with the internet and social media is increasingly essential. It is nearly impossible for a proficient internet user to appreciate the extent of the challenge posed by "opening content in new tabs/windows" for someone much less internet proficient. It is tempting to look to our own experiences to begin to study the experiences of others, but in the case of a technology as inherently heterogeneous as social media, this introspection will necessarily lead scholars astray.

4 Using Facebook Ads to Reach Low Digital Literacy Populations

We used Facebook advertisements to recruit low digital literacy subjects to study the dynamics of online "clickbait." Facebook ads with quota sampling have recently been shown to generate valid measures of public opinion (Zhang et al., 2018), but we were interested in sampling low digital literacy individuals: people who clicked on our eye-catching advertisement. We conducted this experiment to complement a series of experiments on MTurk.

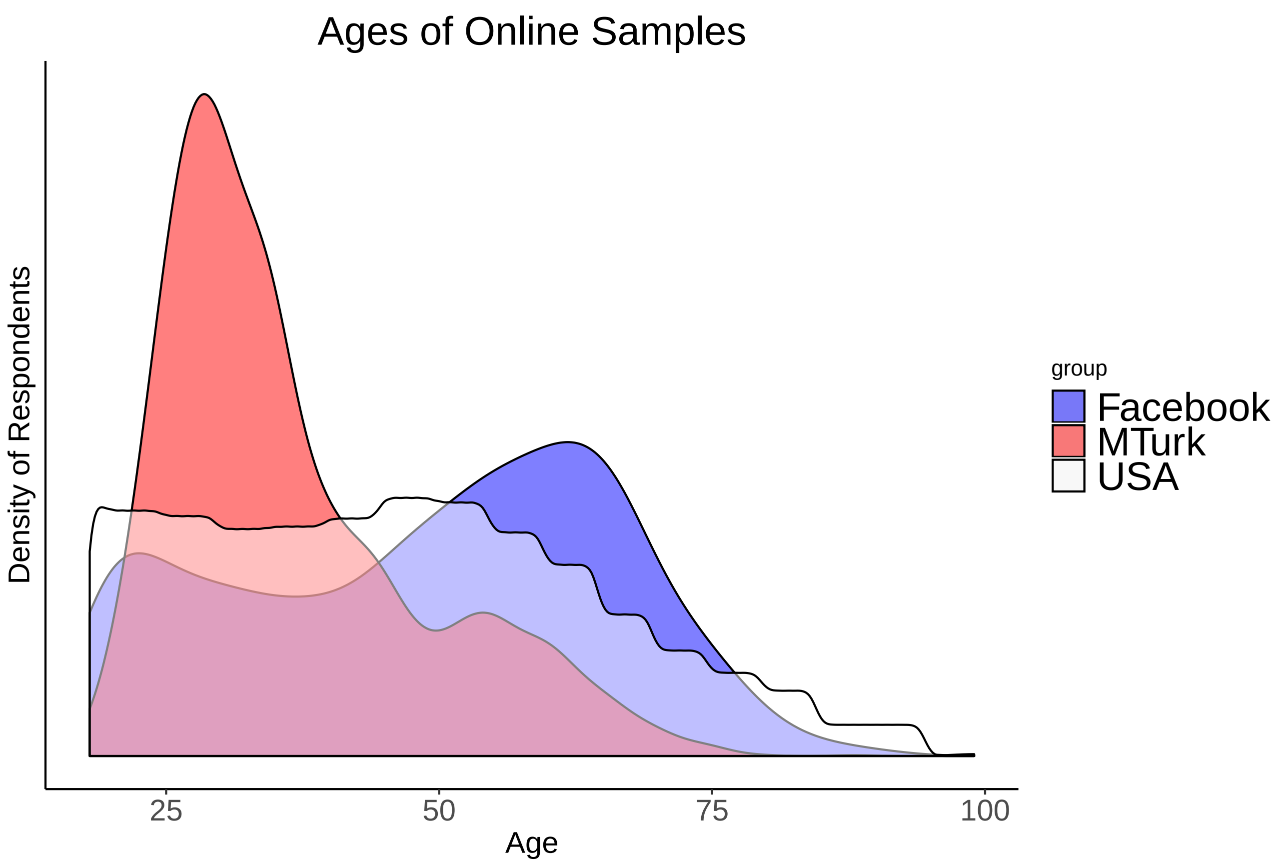
Figure 1 displays the age distributions of the two samples, relative to the 2010 census. The MTurk sample dramatically oversamples adults 24 to 35, and contains exactly 5 people over 75 years old; 2 of these claim to be 99, evidence of unserious responses. In contrast, Facebook oversamples adults between the ages of 50 and 75, and contains a non-trivial number of adults in their 80s.[5] Because the Facebook population is large and the ad targeting well-developed, it is possible to use quota sampling to generate a sample that corresponds to the general population on one or more demographics—and even specific demographic crosstabs—of interest.

Although our initial sample of Facebook users had excellent coverage at all ages, attrition from our survey was non-random. This is evidence that we have encountered an appropriately low digital literacy sample—to such an extent our survey was too difficult for many of them to complete.

This interpretation is supported by Brewer, Morris, and Piper (2016)'s finding that some "barriers, which may seem trivial from a requester's perspective, *significantly*

[5]We cannot be sure whether the age distribution reflects the true rate at which people clicked on ads because there is some uncertainty about the way that the Facebook advertising software operates (Zhang et al., 2018).The rank ordering of propensity to opt into this sample is legitimate, but the true ratio of propensities is opaque.

Figure 1: Age Distributions of Samples: MTurk v Facebook Ads



affected older adults' abilities and time required to complete the tasks" (emphasis in original).

We have three pieces of "forensic" evidence from our survey that this took place. First, we inspect the answers entered into an open response text box asking the respondent's age. Out of 2,803 respondents recruited from MTurk, there were three responses more than two characters long: 999, 999, and 566. Out of 2,467 respondents recruited from Facebook, there were thirty-nine such responses.

Many of these appear to have been due to typos of some kind (eg ",64"), suggesting a lack of digital dexterity. Others, though, indicated the same kind of misunderstanding of the purpose of online surveys described by Brewer, Morris, and Piper (2016), such as "68 yrs. Old. Live. Chicago. With. My. Sister. And. Her. Husband. I am. Wildow".

The mean age of the respondents who entered a two-digit age was 48.8; for those who entered a non-numeric age,[6] it was 62.7, significant at $p < .00001$.

Second, we look at relative attrition rates at different points in the survey. Figure 2 plots attrition at four stages; there is dramatic attrition for the Facebook sample (but not the MTurk sample) at the stage where the survey required clicking on a hyperlinked headline that opened up the news story in a separate tab. At this point in the survey, 6% of the MTurk sample dropped out, compared to 31% of the Facebook sample.

Figure 3 plots the age distributions of subjects based on how much of the survey they completed. The top panel indicates that age is entirely unrelated to attrition stage for the MTurk sample. The bottom panel, however, indicates that the Facebook subjects who completed the entire survey were *much* younger those who did not; those who stopped at the new tab were the oldest of the three.

Our third piece of "forensic" evidence comes from the attention check we embedded in the survey. Attention checks are designed to weed out unserious respondents, but they can also prove confusing to more digitally naive respondents. Our attention check replaced one option in a media choice task with the phrase "Survey taker: always select this option."

Overall, 82% of the MTurk sample "passed" the attention check, compared to just 52% of the Facebook sample. However, some Facebook subjects may have been confused, rather than intentionally providing low-quality responses. Table 1 presents the results of a regression in which the dependent variable is a dummy for whether the respondent stopped when taken to a new tab. The coefficient on the interaction term

[6]Neither sample had anyone over 100 years old.

Figure 2: Relative Attrition Rates

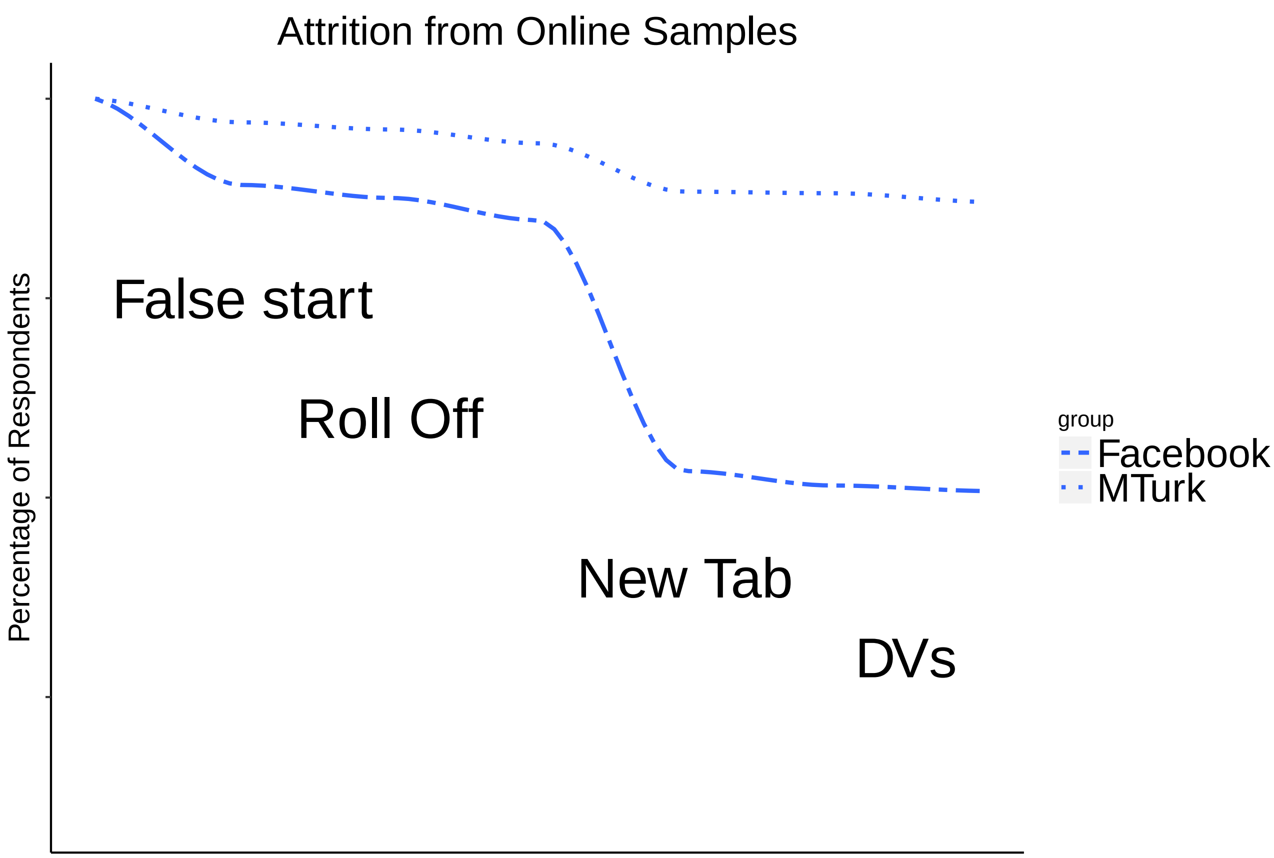
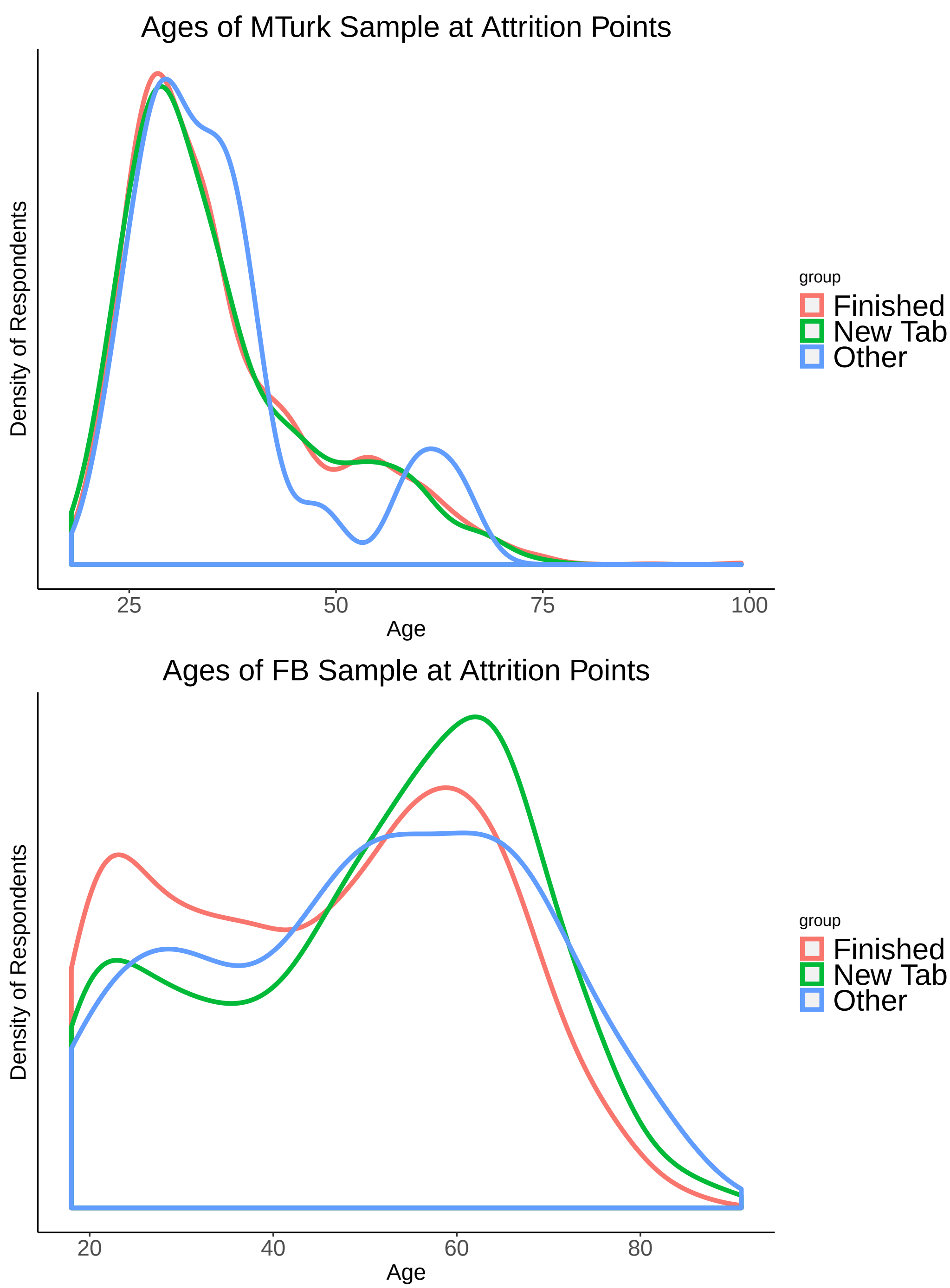


Figure 3: Age Distributions By Attrition Status



between the time spent on the attention check question and a dummy for whether the respondent was in the Facebook sample is positive and highly significant. On the other hand, the time subjects spent on a non-attention check media choice question is unrelated to whether they stopped at the new tab.

Table 1: Facebook Sample Confused by Attention Check Also Confused by New Tab

	Stopped at New Tab
Seconds Spent on Attention Check	-0.002^{*}
	(0.001)
Facebook Sample	0.267^{***}
	(0.028)
Seconds Spent on Standard Choice Question	0.0002
	(0.0003)
Seconds Spent on Attention Check X Facebook Sample	0.004^{***}
	(0.001)
Seconds Spent on Standard Choice Question X Facebook Sample	-0.0001
	(0.0004)
Constant	0.222^{***}
	(0.023)
Observations	3,184
R^2	0.093
Adjusted R^2	0.091

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

In the Facebook sample, both the attention check and the new tab confused subjects. The coefficient on the uninteracted variable for time spent on the attention check indicates that the *opposite* is true for the MTurk sample: those who spent more time on the attention check were *less* likely to stop at the new tab. The attention check worked as intended on the MTurk sample.

5 Conclusion

The immediate argument advanced in this note is that there exists a class of research questions (moderated by age or digital literacy) for which MTurk is an inappropriate source of research subjects. We recommend the continued exploration of Facebook ads as a tool for recruiting subjects from otherwise hard-to-reach populations, of which older or low digital literacy internet users are the most relevant for the study of digital media effects.

However, these populations present novel challenges for researchers; the experience of taking a "standard" online survey may be confusing and overwhelming for digitally naive subjects. As a result, there is both a practical *and* ethical mandate that researchers take care in enhancing the accessibility of online surveys when studying older or low digital literacy populations.

Brewer, Morris, and Piper (2016)'s ethnographic work makes these ethical concerns more emotionally immediate: "These [accessibility] challenges also affected older adults' self-efficacy, with P7 saying, 'I just think I'm not smart enough to do it'; 'I just didn't understand anything they were telling me to do... I'm a complete failure'; and 'I don't even understand the instructions. Is everybody else that does this as dumb as I am?'"

Feeling like "a complete failure" is not a good outcome for participating in an academic study. Our study was conducted online and thus lower stakes than Brewer, Morris, and Piper (2016)'s in-person experiment, but the high attrition suggests that many of our subjects were frustrated, confused or otherwise dissatisfied. Our ignorance led us to design an imperfect survey, which we deeply regret. Our hope is that the lessons we learned will be taken up by other political scientists, and future mistakes avoided.

Practically, we encourage researchers using online survey instruments to make them shorter and less technically challenging to use. This is not a novel point, but it is increasingly urgent when studying populations with low digital literacy. A simple modification would have been to eliminate the need to open and close the tab containing the news story. Another important step is the implementation of small scale, qualitative, pilot studies to ensure that survey instruments are functioning as intended.

On a more theoretical level, we argue that the effects of digital media are far more heterogeneous than any previous form of media. The internet offers an essentially unlimited choice of information sources; these choices *and* their effects are endogenous to an individuals' age and digital literacy. This will only more notable as an increasing number of older people adopt the internet and social media. The “average effect” of digital media is a potentially misleading quantity; based on our study of low digital literacy populations, we encourage researchers to focus on theoretically interesting subpopulations of internet users when studying digital media effects.

References

Barbera, Pablo. 2018. Explaining the Spread of Misinformation on Social Media: Evidence from the 2016 U.S. Presidential Election. In *Symposium: Fake News and the PoliticsofMisinformation*. APSA.

Beech, Hannah, and Saw Nang. 2018. "In Myanmar, a Facebook Blackout Brings More Anger Than a Genocide Charge." *New York Times* August 31, 2018.

Bond, Robert M, Jaime E Settle, Christopher J Fariss, Jason J Jones, and James H Fowler. 2017. "Social endorsement cues and political participation." *Political Communication* 34 (2): 261–281.

Boyd, Danah. 2014. *It's complicated: The social lives of networked teens*. Yale University Press.

Brewer, Robin, Meredith Ringel Morris, and Anne Marie Piper. 2016. Why would anybody do this?: Understanding older adults' motivations and challenges in crowd work. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM pp. 2246–2257.

Coppock, Alexander. 2018. "Generalizing from survey experiments conducted on mechanical Turk: A replication approach." *Political Science Research and Methods* pp. 1–16.

Coppock, Alexander, Thomas J Leeper, and Kevin J Mullinix. 2018. "The Generalizability of Heterogeneous Treatment Effect Estimates Across Samples.".

DiMaggio, Paul, Eszter Hargittai et al. 2001. "From the 'digital divide'to 'digital inequality': Studying Internet use as penetration increases." *Princeton: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University* 4 (1): 4–2.

Guess, Andrew, Jonathan Nagler, and Joshua A. Tucker. 2018. "Who's Clogging Your Facebook Feed? Ideology and Age as Predictors of Fake News Dissemination During the 2016 U.S. Campaign." *Unpublished manuscript* .

Hargittai, Eszter. 2001. "Second-level digital divide: mapping differences in people's online skills." *arXiv preprint cs/0109068* .

Hargittai, Eszter, Anne Marie Piper, and Meredith Ringel Morris. 2018. "From internet access to internet skills: digital inequality among older adults." *Universal Access in the Information Society* pp. 1–10.

Huff, Connor, and Dustin Tingley. 2015. ""Who Are These People?" Evaluating the Demographic Characteristics and Political Preferences of MTurk Survey Respondents." *Research and Politics* 2 (1): 1–12.

Mullinix, Kevin J, Thomas J Leeper, James N Druckman, and Jeremy Freese. 2015. "The generalizability of survey experiments." *Journal of Experimental Political Science* 2 (2): 109–138.

Prensky, Marc. 2001. "Digital natives, digital immigrants part 1." *On the horizon* 9 (5): 1–6.

Smith, Aaron, and Monica Anderson. 2018. *Social Media Use in 2018.* Pew.

Snowberg, Erik, and Leeat Yariv. 2018. "Testing the Waters: Behavior across Participant Pools.".

Zhang, Baobao, Matto Mildenberger, Peter D. Howe, Jennifer Marlon, Seth Rosenthal, and Anthony Leiserowitz. 2018. "Quota Sampling Using Facebook Advertisements." *Political Science Research and Methods* pp. 1–16.