Insufficient Variation in Treatment Moderators: The Limitations of Mechanical Turk for Studying Digital Media Effects

Kevin Munger, Mario Luca, Jonathan Nagler, Joshua Tucker

Penn State University and Princeton University

December 7, 2018

Overview

- This is an in-depth discussion of a series of experiments that "failed"

Overview

- This is an in-depth discussion of a series of experiments that "failed"
- Null, precise treatment effects

Overview

- This is an in-depth discussion of a series of experiments that "failed"
- Null, precise treatment effects
- Walk through our experiences as researchers

Overview

- This is an in-depth discussion of a series of experiments that "failed"
- Null, precise treatment effects
- Walk through our experiences as researchers
- Engage in some "methodological self-criticism"

Overview

- This is an in-depth discussion of a series of experiments that "failed"
- Null, precise treatment effects
- Walk through our experiences as researchers
- Engage in some "methodological self-criticism"
- Discuss the field and propose improvements for future studies

Clickbait

- Clickbait is an understudied but prominent form of online media

Clickbait

- Clickbait is an understudied but prominent form of online media
- What kind of people are the most likely to consume clickbait media?

Clickbait

- Clickbait is an understudied but prominent form of online media
- What kind of people are the most likely to consume clickbait media?
- What are the implications of reading clickbait headlines for affective polarization, trust in media and information retention?

Clickbait

- Clickbait is an understudied but prominent form of online media
- What kind of people are the most likely to consume clickbait media?
- What are the implications of reading clickbait headlines for affective polarization, trust in media and information retention?
- How can we find the relevant population: want to estimate the effect of clickbait on people who actually read clickbait?

Clickbait

- Clickbait is an understudied but prominent form of online media
- What kind of people are the most likely to consume clickbait media?
- What are the implications of reading clickbait headlines for affective polarization, trust in media and information retention?
- How can we find the relevant population: want to estimate the effect of clickbait on people who actually read clickbait?

What is clickbait?

- “Clickbait” is a new term for an old phenomenon.

What is clickbait?

- "Clickbait" is a new term for an old phenomenon.
- Media companies' strategy always determined by technological, political, regulatory contexts

What is clickbait?

- "Clickbait" is a new term for an old phenomenon.
- Media companies' strategy always determined by technological, political, regulatory contexts
- New technology lowers cost of news production/distribution $\rightsquigarrow$ new entrants competing for attention

Classic Clickbait

$50,000 REWARD.—WHO DESTROYED THE MAINE?—$50,000 REWARD.

EDITION FOR GREATER NEW YORK.

NEW YORK JOURNAL

AND ADVERTISER.

NO. 3,372. NEW YORK, THURSDAY, FEBRUARY 17, 1898.—16 PAGES. PRICE ONE CENT

DESTRUCTION OF THE WAR SHIP MAINE WAS THE WORK OF AN ENEMY

$50,000!

$50,000 REWARD!

For the Detection of the Perpetrator of the Maine Outrage!

Assistant Secretary Roosevelt Convinced the Explosion of the War Ship Was Not an Accident.

The Journal Offers $50,000 Reward for the Conviction of the Criminals Who Sent 258 American Sailors to Their Death. Naval Officers Unanimous That the Ship Was Destroyed on Purpose.

$50,000!

$50,000 REWARD!

For the Detection of the Perpetrator of the Maine Outrage!

Experimental Design

- Online survey experiment

Experimental Design

- Online survey experiment
- Ask respondents to choose which story they'd like to read

Experimental Design

- Online survey experiment

- Ask respondents to choose which story they'd like to read
  - Estimate individual-level Preference for Clickbait

Experimental Design

- Online survey experiment

- Ask respondents to choose which story they'd like to read
  - Estimate individual-level Preference for Clickbait
  - Model how Preference for Clickbait varies among different kinds of people

Experimental Design

- Online survey experiment

- Ask respondents to choose which story they'd like to read
  - Estimate individual-level Preference for Clickbait
  - Model how Preference for Clickbait varies among different kinds of people

- Randomly assign respondents to one of four different headlines, keeping the story constant

Experimental Design

- Online survey experiment

- Ask respondents to choose which story they'd like to read
  - Estimate individual-level Preference for Clickbait
  - Model how Preference for Clickbait varies among different kinds of people

- Randomly assign respondents to one of four different headlines, keeping the story constant
  - Emotional Clickbait vs traditional headline
  - Democrat or Republican headline

Experimental Design

- Online survey experiment

- Ask respondents to choose which story they'd like to read
  - Estimate individual-level Preference for Clickbait
  - Model how Preference for Clickbait varies among different kinds of people

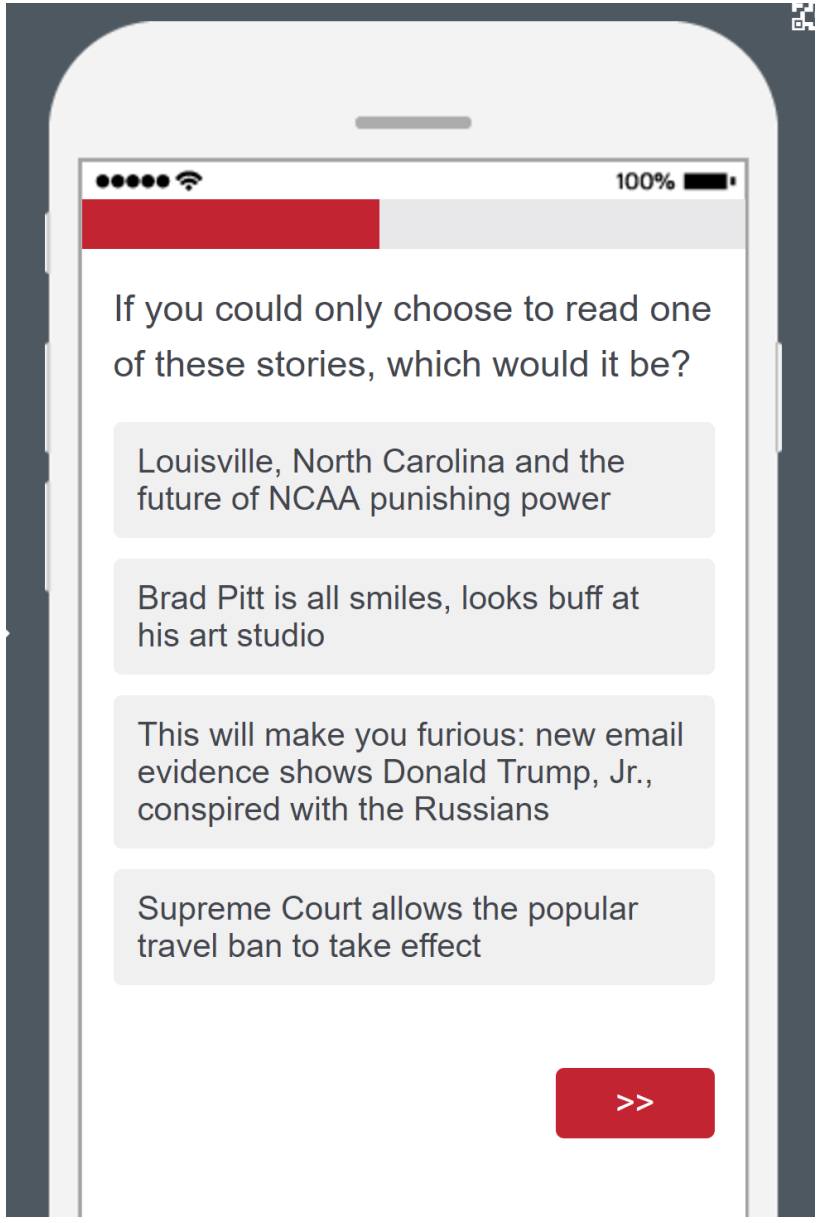
- Randomly assign respondents to one of four different headlines, keeping the story constant
  - Emotional Clickbait vs traditional headline
  - Democrat or Republican headline

- Ask affect polarization, trust in media and information retention questions

Experimental Design

- Online survey experiment
- Ask respondents to choose which story they'd like to read
  - Estimate individual-level Preference for Clickbait
  - Model how Preference for Clickbait varies among different kinds of people
- Randomly assign respondents to one of four different headlines, keeping the story constant
  - Emotional Clickbait vs traditional headline
  - Democrat or Republican headline
- Ask affect polarization, trust in media and information retention questions
- After the pilot, we pre-registered the R code we used to analyze all experimental results

Estimating "Preference for Clickbait"



Samples—Will Return to This

- Conducted two experiments using samples recruited from Amazon's Mechanical Turk, oversampling Republicans to ensure a balanced sample

Samples—Will Return to This

- Conducted two experiments using samples recruited from Amazon's Mechanical Turk, oversampling Republicans to ensure a balanced sample
- Only the first survey contained the Preference for Clickbait battery

Samples—Will Return to This

- Conducted two experiments using samples recruited from Amazon's Mechanical Turk, oversampling Republicans to ensure a balanced sample
- Only the first survey contained the Preference for Clickbait battery
- Identical survey on a sample of subjects collected via Facebook ads

	Preference for Clickbait		Preference for Republican	
	MTurk	FB	MTurk	FB
Facebook use	0.003	0.015^{***}	0.004	0.007
	(0.003)	(0.006)	(0.003)	(0.007)
Twitter use	0.003	0.002	-0.001	-0.001
	(0.003)	(0.002)	(0.003)	(0.002)
Internet use	0.009	-0.009^{*}	0.002	0.009
	(0.008)	(0.005)	(0.009)	(0.006)
Age	0.001^{*}	0.001^{***}	0.001	-0.0001
	(0.0005)	(0.0003)	(0.001)	(0.0003)
Education	-0.005	-0.014^{***}	-0.001	-0.007
	(0.007)	(0.004)	(0.008)	(0.005)
Offline news consumption	0.009^{**}	0.001	-0.0001	-0.001
	(0.004)	(0.002)	(0.004)	(0.002)
Offline news consumption	0.006	0.005^{*}	-0.002	-0.012^{***}
	(0.005)	(0.003)	(0.005)	(0.003)
Democrat	-0.052^{***}	-0.020^{*}	-0.129^{***}	-0.096^{***}
	(0.016)	(0.010)	(0.018)	(0.012)
Lean Democrat	-0.032^{*}	-0.032^{**}	-0.082^{***}	-0.090^{***}
	(0.019)	(0.015)	(0.021)	(0.018)
Lean Republican	0.067^{***}	0.012	0.112^{***}	0.103^{***}
	(0.018)	(0.017)	(0.020)	(0.020)
Republican	0.036^{**}	0.021	0.135^{***}	0.165^{***}
	(0.018)	(0.014)	(0.020)	(0.016)

Experimental Results

Null Results from First MTurk Study

- Tried again: changed the topic of the articles, added a placebo condition

Null Results from Second MTurk Study

- Tried again: shortened the survey, removed "preference for clickbait" questionnaire which could dampen treatment effects

Null Results from Second MTurk Study

- Tried again: shortened the survey, removed "preference for clickbait" questionnaire which could dampen treatment effects
- Again, null results!

Is Mturk the problem?

Is Mturk the problem?

- Many classic (non-digital) experiments replicate on MTurk (Coppock, 2018)

Is Mturk the problem?

- Many classic (non-digital) experiments replicate on MTurk (Coppock, 2018)
- Econ-style experiments also largely replicate on MTurk compared to students or a nationally representative sample (Snowberg and Yariv, 2018)

Is Mturk the problem?

- Many classic (non-digital) experiments replicate on MTurk (Coppock, 2018)
- Econ-style experiments also largely replicate on MTurk compared to students or a nationally representative sample (Snowberg and Yariv, 2018)
- Massive replication of 27 "framing, priming and information survey experiments" finds replication of both ATEs and CATEs (Coppock, Leeper, and Mullinix, 2018)

Is Mturk the problem?

- Many classic (non-digital) experiments replicate on MTurk (Coppock, 2018)
- Econ-style experiments also largely replicate on MTurk compared to students or a nationally representative sample (Snowberg and Yariv, 2018)
- Massive replication of 27 "framing, priming and information survey experiments" finds replication of both ATEs and CATEs (Coppock, Leeper, and Mullinix, 2018)
- But this is only for the *already-theorized* treatment moderators

Is Mturk the problem?

- Many classic (non-digital) experiments replicate on MTurk (Coppock, 2018)
- Econ-style experiments also largely replicate on MTurk compared to students or a nationally representative sample (Snowberg and Yariv, 2018)
- Massive replication of 27 "framing, priming and information survey experiments" finds replication of both ATEs and CATEs (Coppock, Leeper, and Mullinix, 2018)
- But this is only for the *already-theorized* treatment moderators
- MTurk users are **all** above a certain threshold of digital literacy

Clickbait Effects on the Clickers

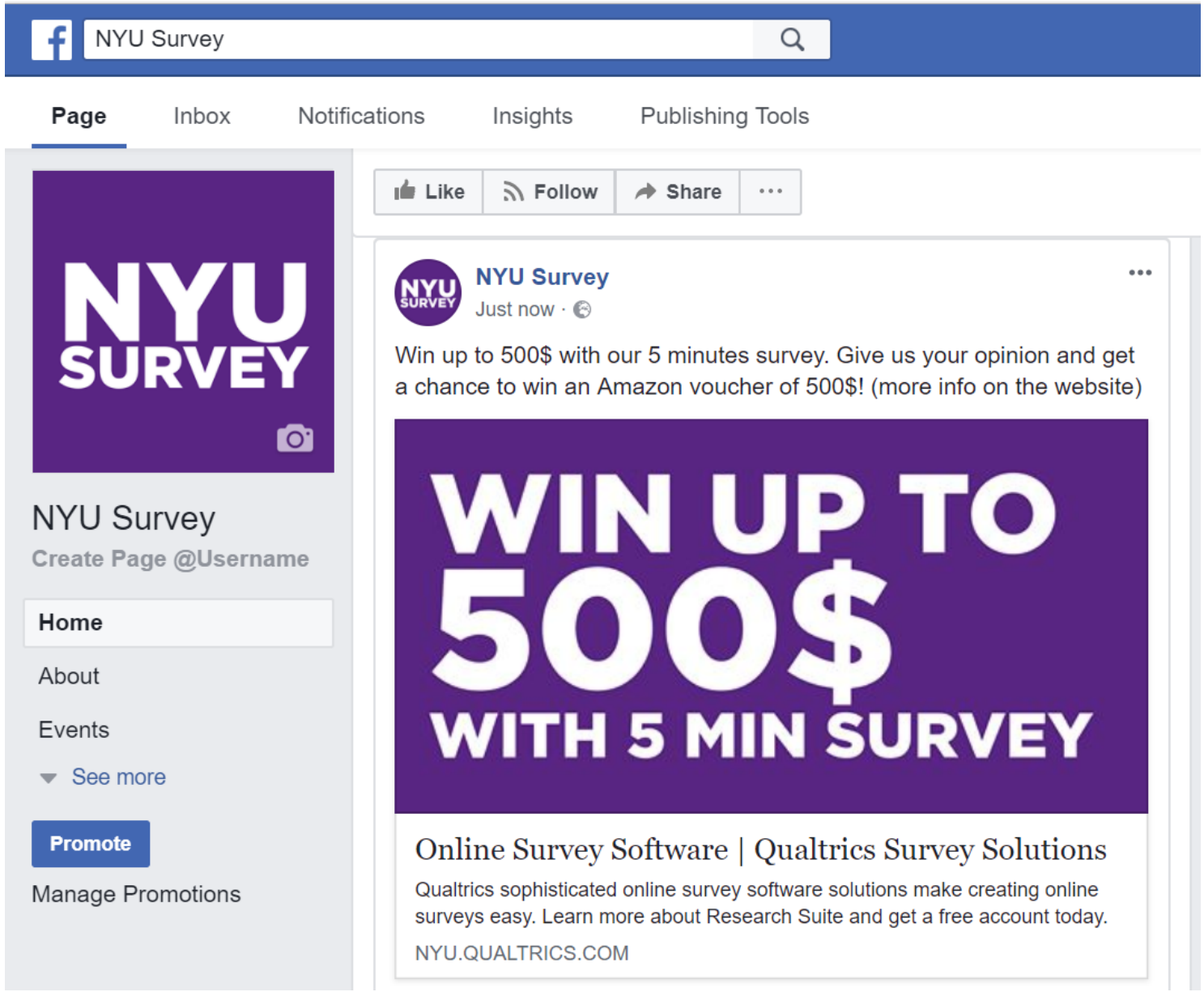
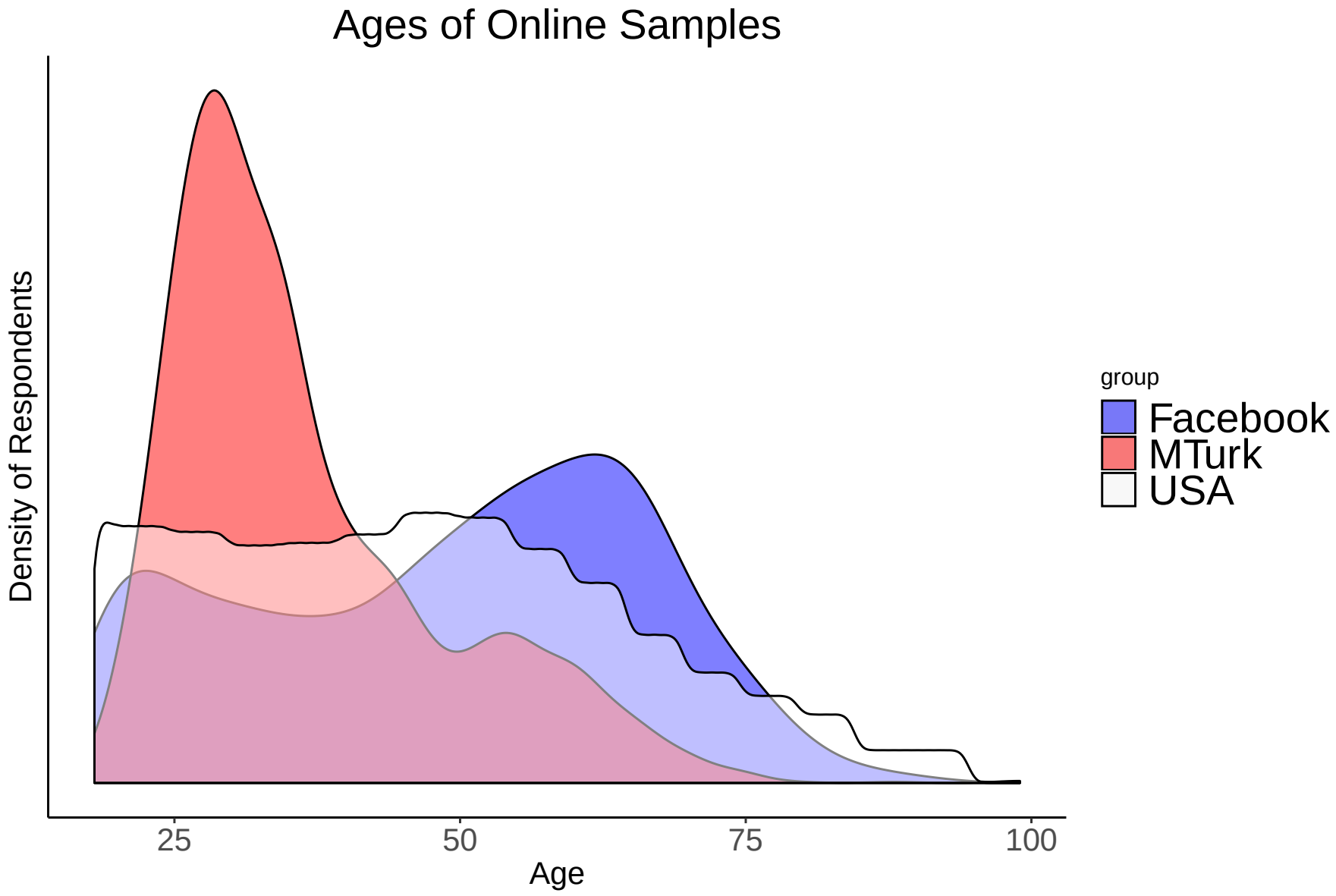


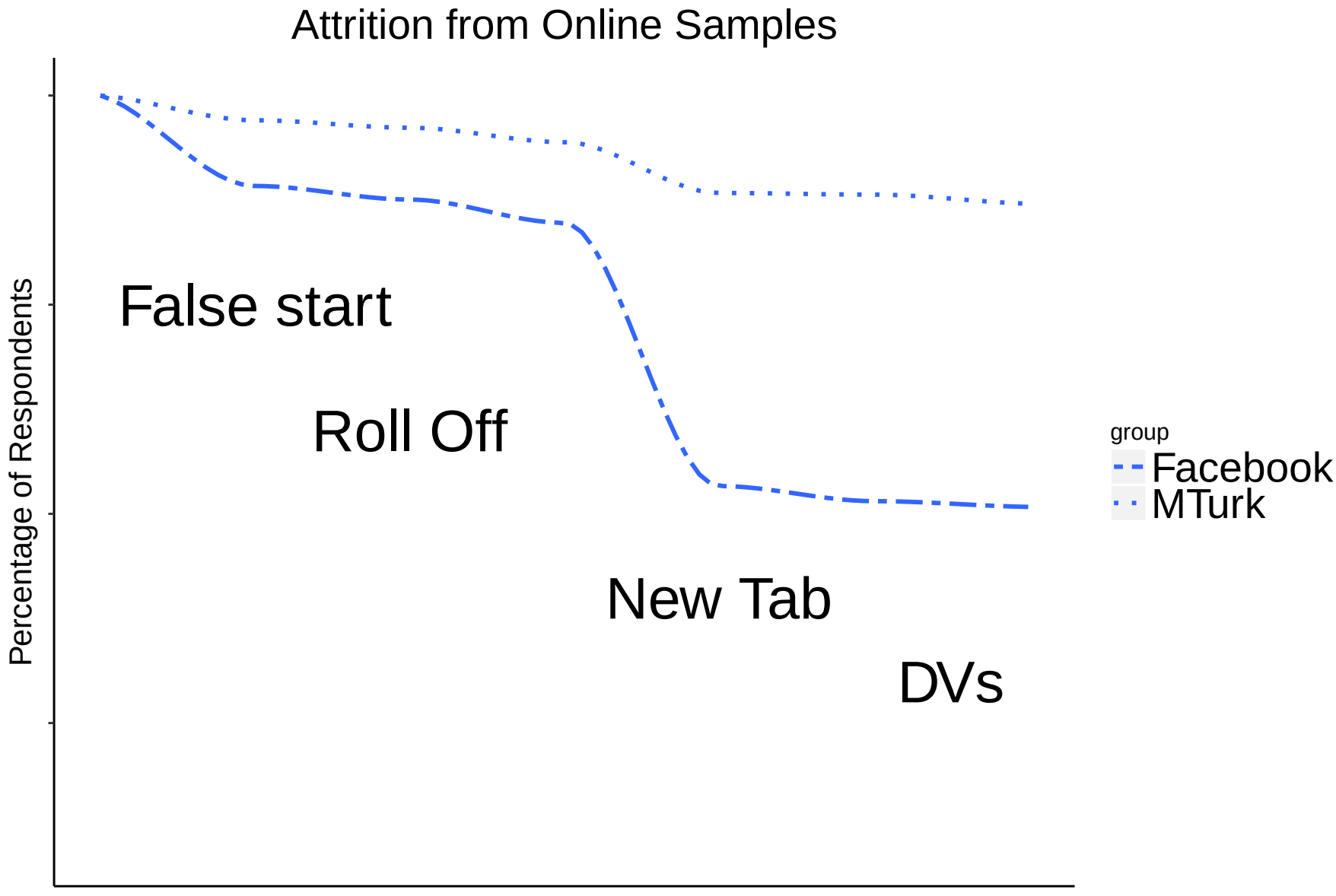
Table: Summary Statistics of MTurk and Facebook Samples

	MTurk	Facebook
% Female	46%	75%
Mean Age	37	49
75th Percentile Age	43	63
% Finished College	58%	42%
% Republican	33%	21%
% Independent	29%	28%
% Internet $>$ 1/day	96%	93%
% Facebook $>$ 1/day	52%	90%
N	1,903	2,382


Ages of Online Samples
Density of Respondents
25
50
75
100
Age
group
Facebook
MTurk
USA

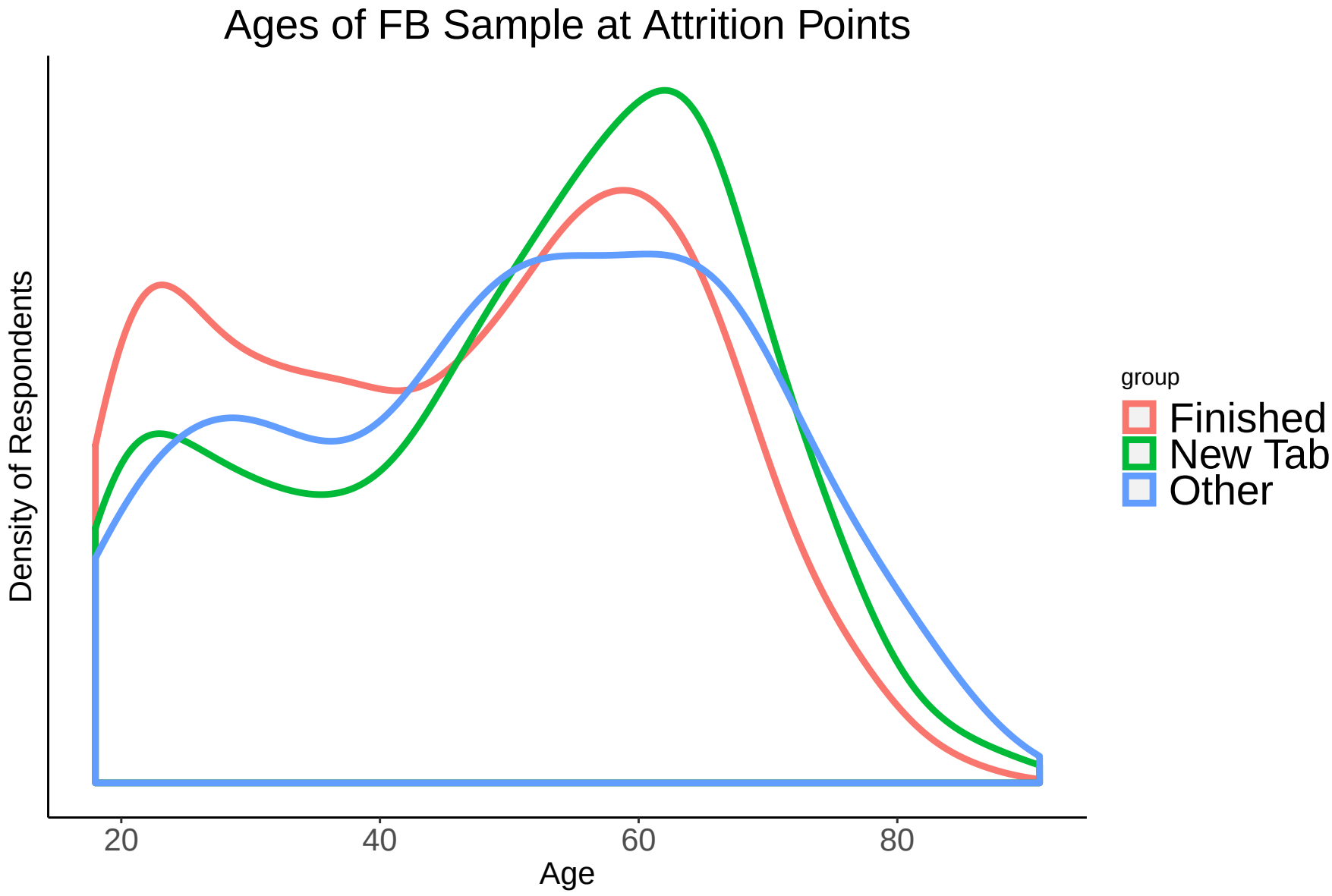
Null Results from the FB Study

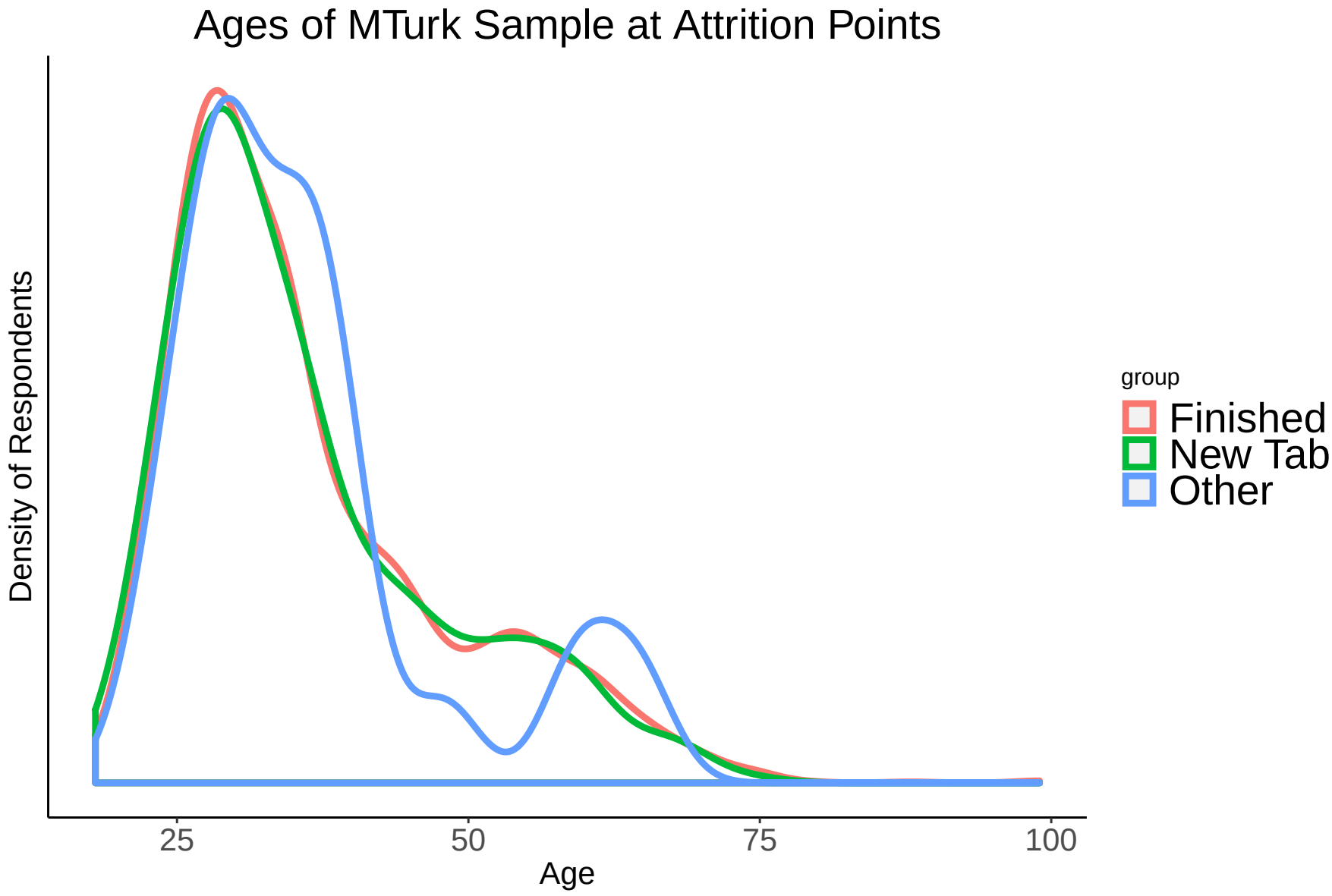
- We got the right sample and didn't find results


Attrition from Online Samples
Percentage of Respondents
False start
Roll Off
New Tab
DVs
group
Facebook
MTurk

Null Results from the FB Study

- ~~We got the right sample and didn't find results~~ Attrition was non-random and covaried with demographics of interest


Ages of FB Sample at Attrition Points
Density of Respondents
20
40
60
80
Age
group
Finished
New Tab
Other

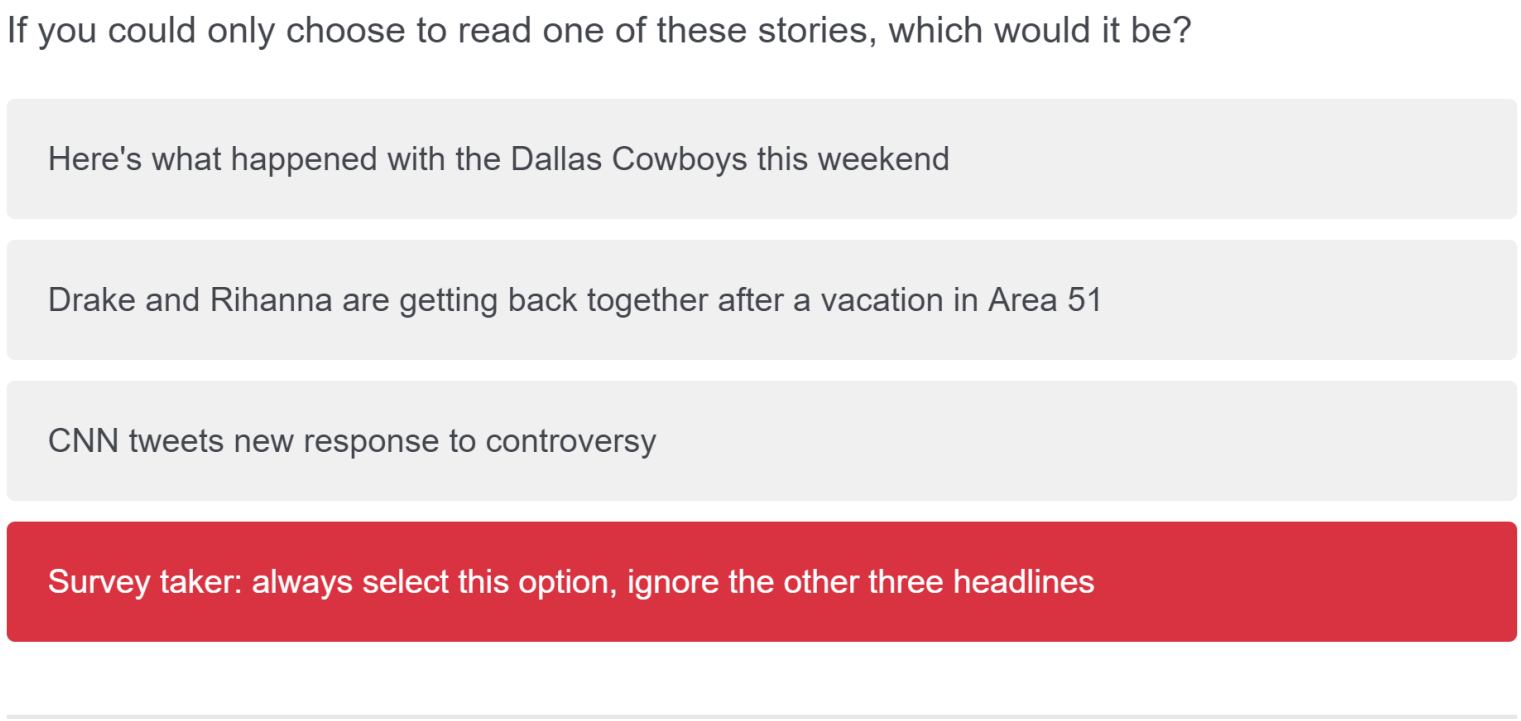

Ages of MTurk Sample at Attrition Points
Density of Respondents
25
50
75
100
Age
group
Finished
New Tab
Other

Examine Predictors of Stopping at New Tab

- Combine the data, run a fully interacted model to look at differential effects in the two samples

Effect of Age on Stopping at New Tab: MTurk v Facebook

"Attention Checks" With Digitally Naive Populations

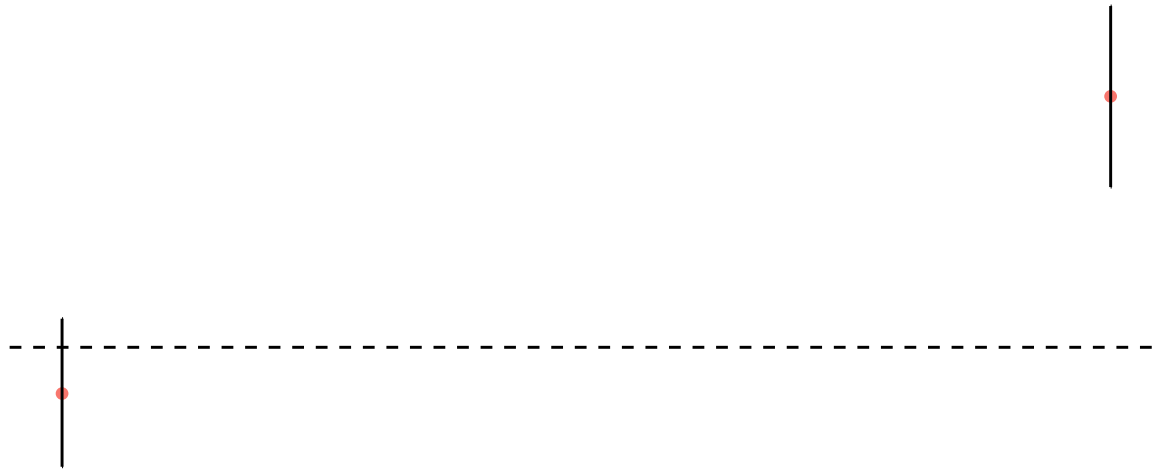


Passed attention check MTurk: 82%

Passed attention check FB: 57%

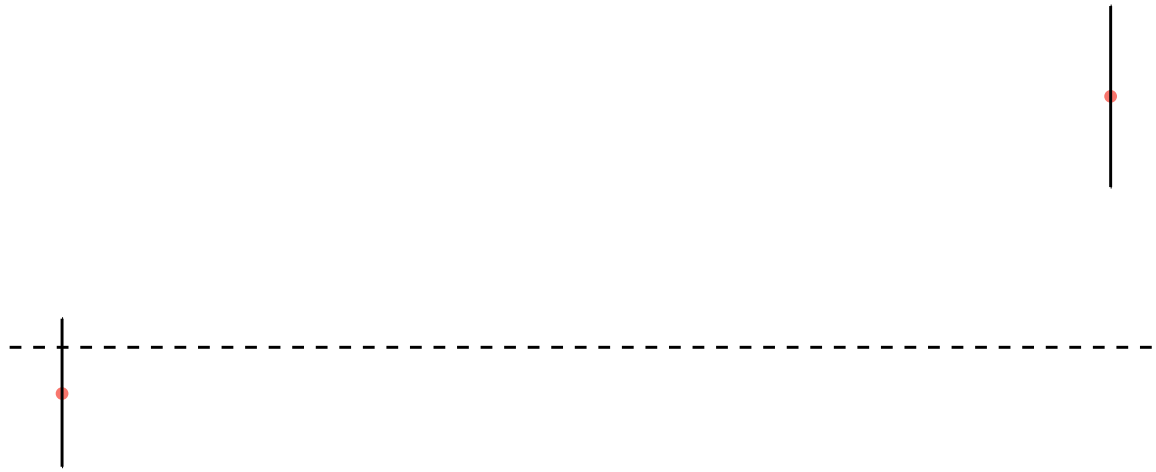
Time Spent on Headline Choice Sets

Attention Check Time on Stopping at New Tab: MTurk v Facebook

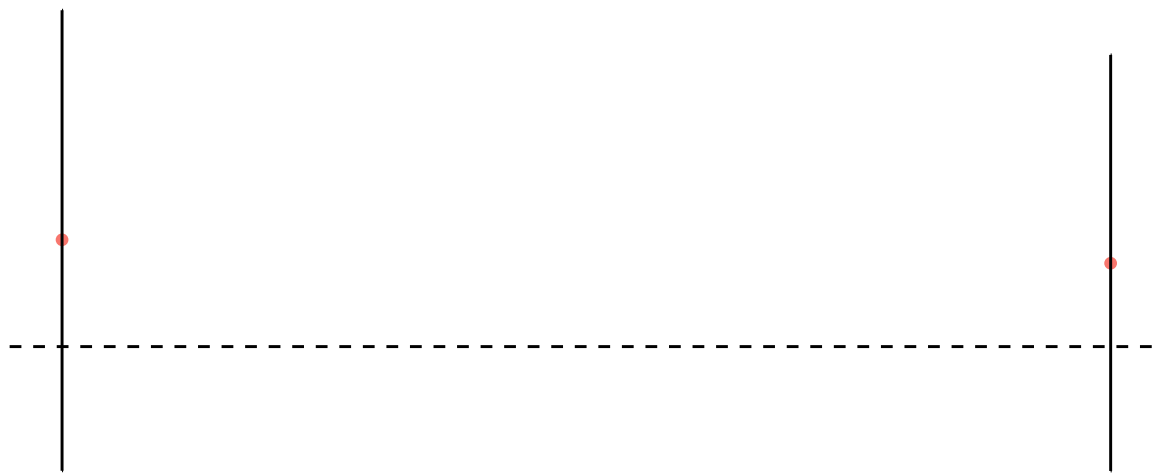


Time Spent on Headline Choice Sets

Attention Check Time on Stopping at New Tab: MTurk v Facebook



Placebo Choice Time on Stopping at New Tab: MTurk v Facebook



Attention Checks

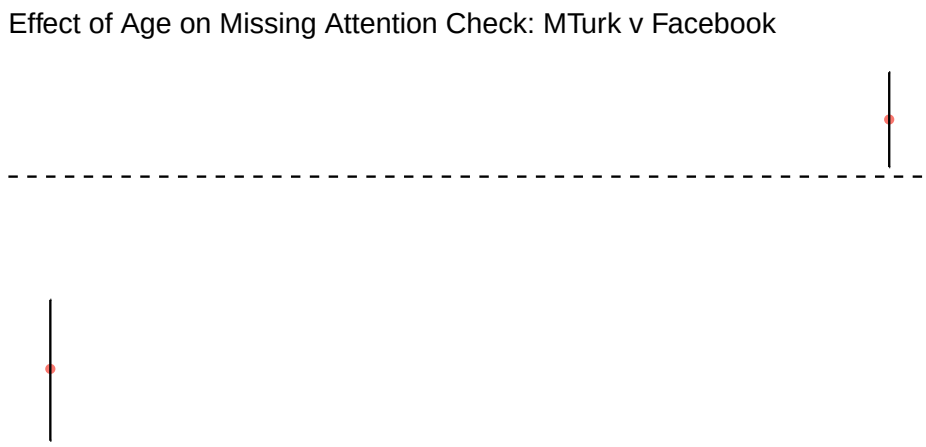
- Combine the data, run a fully interacted model to predict missing the attention check

Attention Checks

- Combine the data, run a fully interacted model to predict missing the attention check
- Differential effects (similar to above)

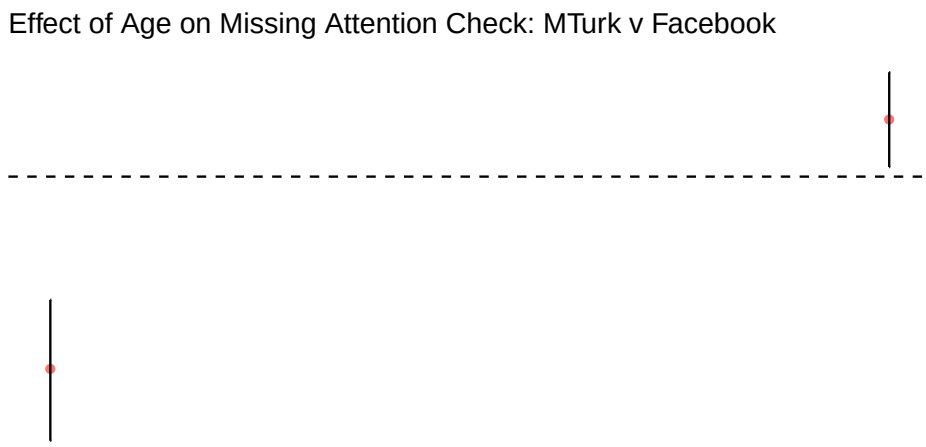
Attention Checks

- Combine the data, run a fully interacted model to predict missing the attention check
- Differential effects (similar to above)



Attention Checks

- Combine the data, run a fully interacted model to predict missing the attention check
- Differential effects (similar to above)



Non-Numeric Ages Entered Into Text Box: Digital Dexterity

- Open-response question: “How old are you”

Non-Numeric Ages Entered Into Text Box: Digital Dexterity

- Open-response question: "How old are you"
- All but 3 MTurkers entered a two-digit response (555, 999, 999)

Non-Numeric Ages Entered Into Text Box: Digital Dexterity

- Open-response question: "How old are you"
- All but 3 MTurkers entered a two-digit response (555, 999, 999)
- 39 Facebookers entered [eg]:

Non-Numeric Ages Entered Into Text Box: Digital Dexterity

- Open-response question: "How old are you"
- All but 3 MTurkers entered a two-digit response (555, 999, 999)
- 39 Facebookers entered [eg]:
  - Seventy one years

Non-Numeric Ages Entered Into Text Box: Digital Dexterity

- Open-response question: "How old are you"
- All but 3 MTurkers entered a two-digit response (555, 999, 999)
- 39 Facebookers entered [eg]:
  - Seventy one years
  - 78 and not senile.

Non-Numeric Ages Entered Into Text Box: Digital Dexterity

- Open-response question: "How old are you"
- All but 3 MTurkers entered a two-digit response (555, 999, 999)
- 39 Facebookers entered [eg]:
  - Seventy one years
  - 78 and not senile.
  - 68 yrs. Old. Live. Chicago. With. My. Sister. And. Her. Husband. I am. Wildow

Non-Numeric Ages Entered Into Text Box: Digital Dexterity

- Open-response question: "How old are you"
- All but 3 MTurkers entered a two-digit response (555, 999, 999)
- 39 Facebookers entered [eg]:
  - Seventy one years
  - 78 and not senile.
  - 68 yrs. Old. Live. Chicago. With. My. Sister. And. Her. Husband. I am. Wildow
  - ,67

Non-Numeric Ages Entered Into Text Box: Digital Dexterity

- Open-response question: "How old are you"
- All but 3 MTurkers entered a two-digit response (555, 999, 999)
- 39 Facebookers entered [eg]:
  - Seventy one years
  - 78 and not senile.
  - 68 yrs. Old. Live. Chicago. With. My. Sister. And. Her. Husband. I am. Wildow
  - ,67
- Average ages: 45 and 63

Older People Using Mechanical Turk

- Qualitative study of older, non-Mturk users: they can't do basic tasks on MTurk (Brewer, Morris, and Piper, 2016)

Older People Using Mechanical Turk

- Qualitative study of older, non-Mturk users: they can't do basic tasks on MTurk (Brewer, Morris, and Piper, 2016)
- Preliminary survey: biggest barrier to crowdwork is not knowing what crowdwork is

Older People Using Mechanical Turk

- Qualitative study of older, non-Mturk users: they can't do basic tasks on MTurk (Brewer, Morris, and Piper, 2016)
- Preliminary survey: biggest barrier to crowdwork is not knowing what crowdwork is
- In-depth study of non-crowdworkers encouraged to sign up and complete basic tasks

Older People Using Mechanical Turk

- Qualitative study of older, non-Mturk users: they can't do basic tasks on MTurk (Brewer, Morris, and Piper, 2016)
- Preliminary survey: biggest barrier to crowdwork is not knowing what crowdwork is
- In-depth study of non-crowdworkers encouraged to sign up and complete basic tasks
- Modal respondent reported having used the internet for 10 years or more

Older People Using Mechanical Turk

many participants were not familiar or comfortable with opening content in new tabs/windows....'How do I get back to the instructions? (P7)'....P3 explained: 'There's too many things to remember all at once...One of my complaints about some things on a computer is that, you know, if there's a bunch of instructions or stuff to know and you have to open up a box and then if you go back to what you're working on the box is gone, and you can't just look up [sic] and reference it. (Brewer, Morris and Piper, 2016)

Older People Using Mechanical Turk

These barriers, which may seem trivial from a requester's perspective, significantly affected older adults' abilities and time required to complete the tasks. These 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 dont even understand the instructions. Is everybody else that does this as dumb as I am?' (Brewer, Morris and Piper, 2016)

"Reflexivity" (Bourdieu)

- The intuitions of social scientists are not merely insufficient; they're *actively misleading*

"Reflexivity" (Bourdieu)

- The intuitions of social scientists are not merely insufficient; they're *actively misleading*
- Models that describe us *don't* describe most people

"Reflexivity" (Bourdieu)

- The intuitions of social scientists are not merely insufficient; they're *actively misleading*
- Models that describe us *don't* describe most people
- Twitter bot experiments—isn't it obvious that these are bots?

"Reflexivity" (Bourdieu)

- The intuitions of social scientists are not merely insufficient; they're *actively misleading*
- Models that describe us *don't* describe most people
- Twitter bot experiments—isn't it obvious that these are bots?
- *Not to non-social scientists*

"Reflexivity": Online Echo Chambers

- Considerable energy has been spent investigating the phenomenon of echo chambers

"Reflexivity": Online Echo Chambers

- Considerable energy has been spent investigating the phenomenon of echo chambers
- They don't exist...except among users in specialized (partisan or professional) networks

"Reflexivity": Online Echo Chambers

- Considerable energy has been spent investigating the phenomenon of echo chambers
- They don't exist...except among users in specialized (partisan or professional) networks
  - Including, of course, social scientists and journalists

"Reflexivity": Online Echo Chambers

- Considerable energy has been spent investigating the phenomenon of echo chambers
- They don't exist...except among users in specialized (partisan or professional) networks
  - Including, of course, social scientists and journalists
- Science works: empirical consensus falsified the theory of ubiquitous echo chambers

"Reflexivity": Online Echo Chambers

- Considerable energy has been spent investigating the phenomenon of echo chambers
- They don't exist...except among users in specialized (partisan or professional) networks
  - Including, of course, social scientists and journalists
- Science works: empirical consensus falsified the theory of ubiquitous echo chambers
- But the supply of social science research is inelastic, so there are serious opportunity costs

Costs of the Bourdieusian Scholastic View

	Online Echo Chambers	Fake News
Experienced by us	YES	NO

Costs of the Bourdieusian Scholastic View

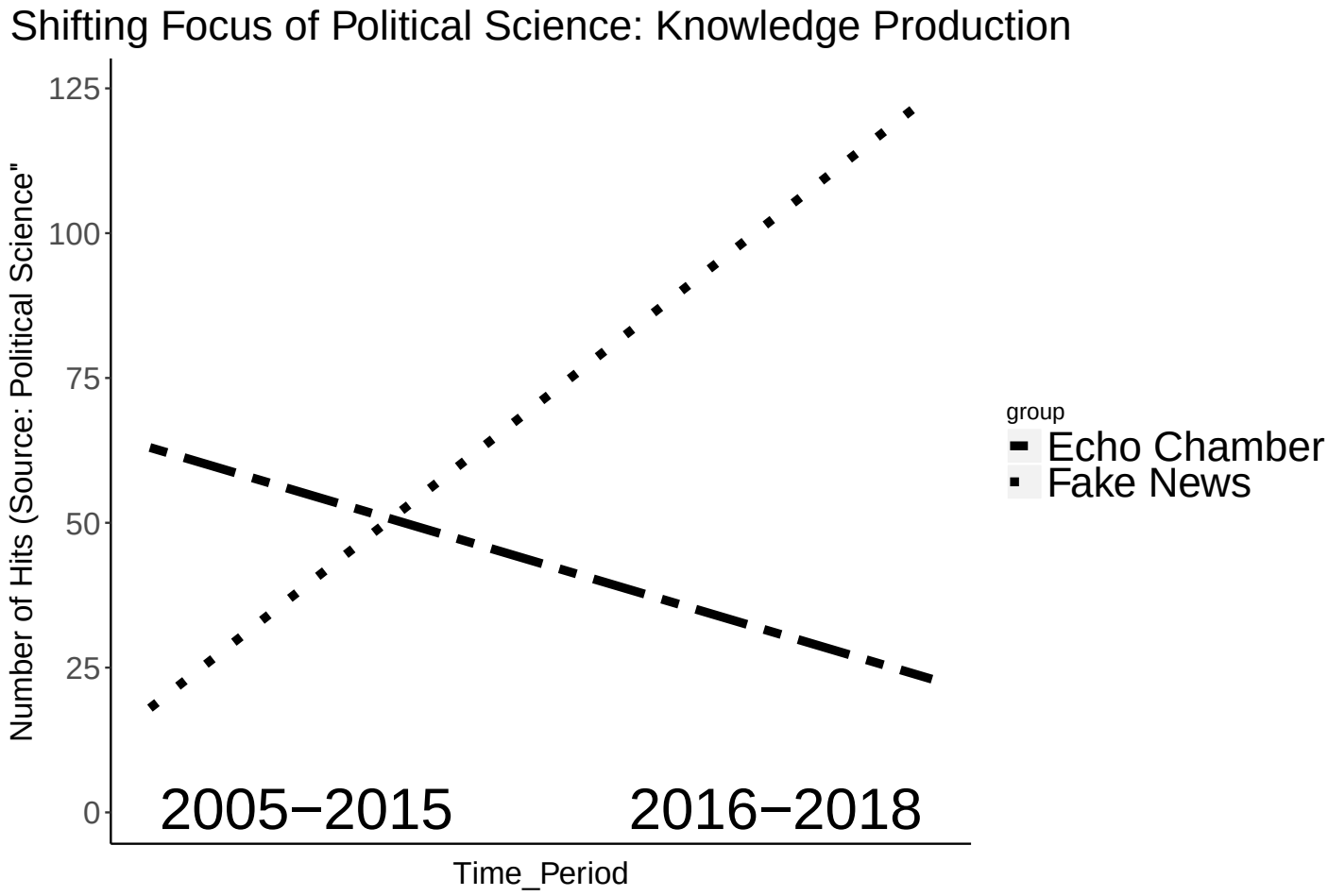
	Online Echo Chambers	Fake News
Experienced by us	YES	NO
Experienced on average	NO	NO

Costs of the Bourdieusian Scholastic View

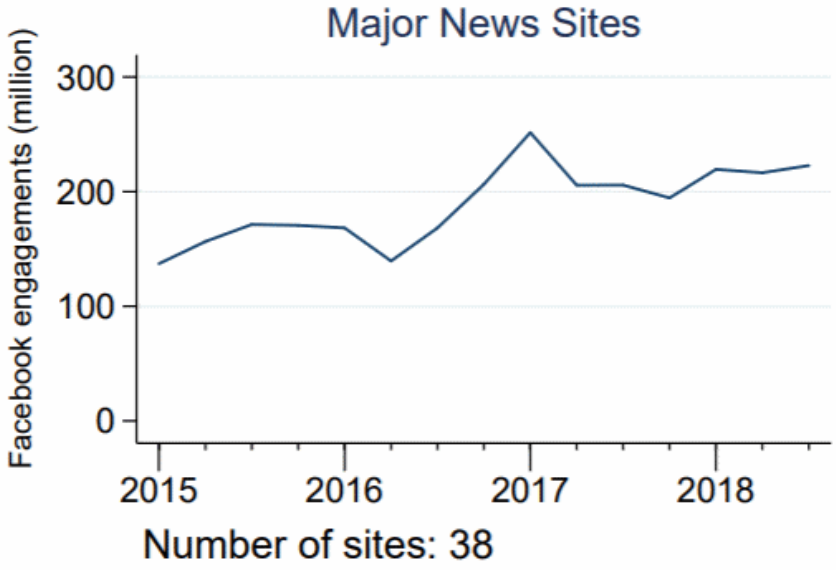
	Online Echo Chambers	Fake News
Experienced by us	YES	NO
Experienced on average	NO	NO
Specific sub-populations	YES	YES

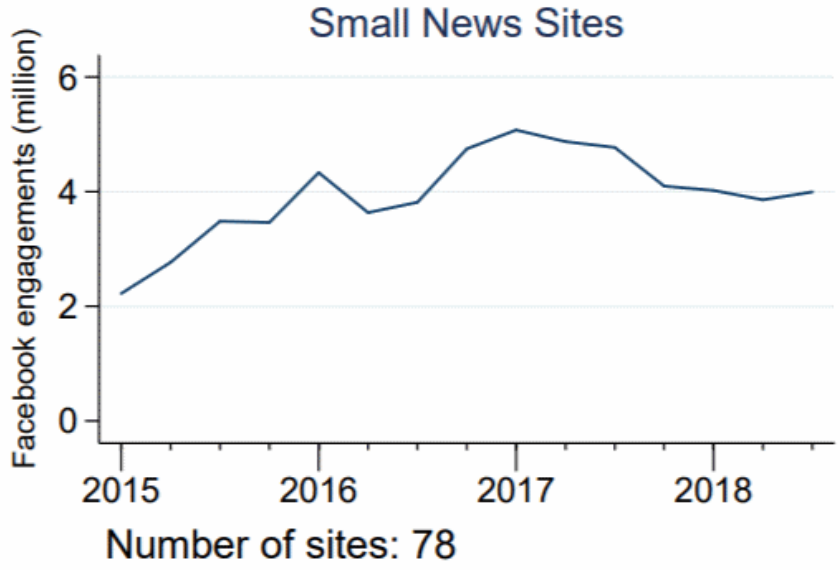
Costs of the Bourdieusian Scholastic View

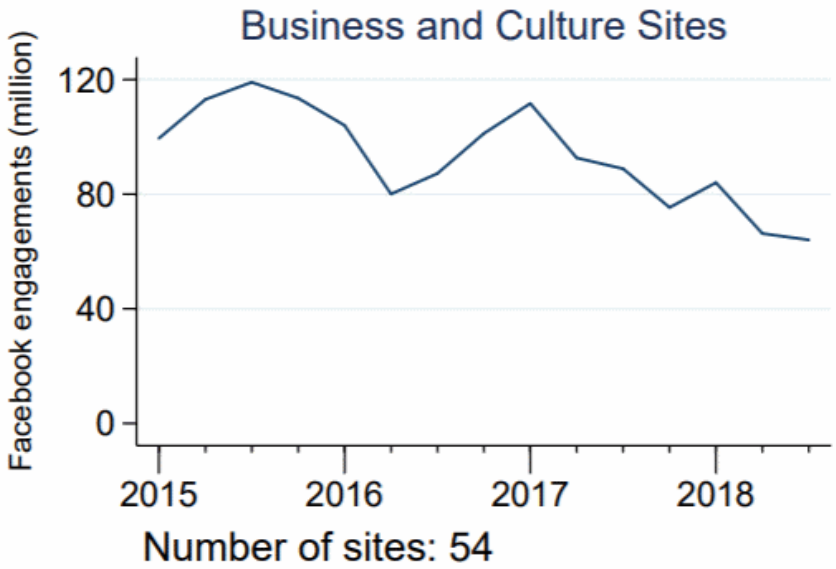
	Online Echo Chambers	Fake News
Experienced by us	YES	NO
Experienced on average	NO	NO
Specific sub-populations	YES	YES
Studied by us	TOO MUCH	TOO LATE

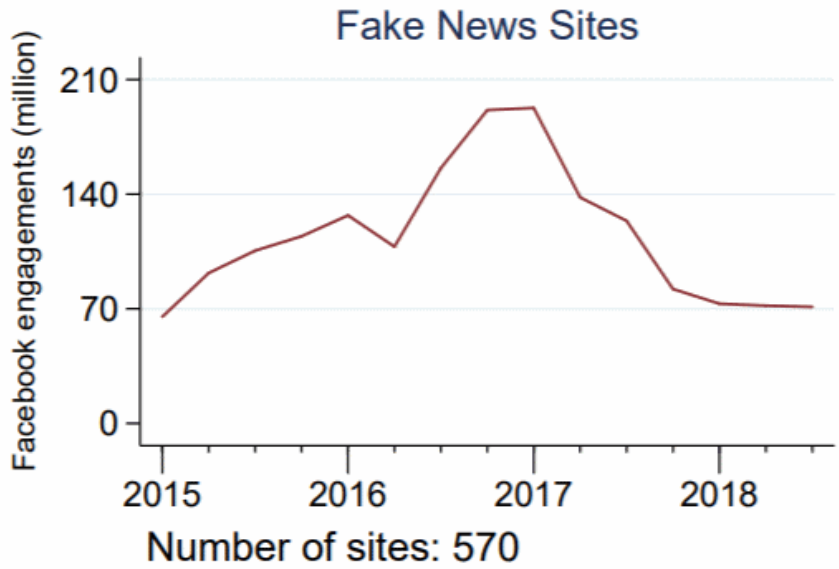

Shifting Focus of Political Science: Knowledge Production
Number of Hits (Source: Political Science"
125
100
75
50
25
0
2005−2015
2016−2018
Time_Period
group
Echo Chamber
Fake News

Panel A: Facebook Engagements









Generalizability

- "The Generalizability of Survey Experiments" (Mullinix et al., 2015):

Generalizability

- "The Generalizability of Survey Experiments" (Mullinix et al., 2015):
  - General point: survey experiments generalize...

Generalizability

- "The Generalizability of Survey Experiments" (Mullinix et al., 2015):
  - General point: survey experiments generalize...
  - "some convenience samples would be inappropriate such as a student sample where a moderator is age"

Generalizability

- "The Generalizability of Survey Experiments" (Mullinix et al., 2015):
  - General point: survey experiments generalize...
  - "some convenience samples would be inappropriate such as a student sample where a moderator is age"

- **Any sample with a hard digital literacy cutoff is inappropriate for making generalizations about online behaviors**

Generalizability

- "The Generalizability of Survey Experiments" (Mullinix et al., 2015):
  - General point: survey experiments generalize...
  - "some convenience samples would be inappropriate such as a student sample where a moderator is age"

- **Any sample with a hard digital literacy cutoff is inappropriate for making generalizations about online behaviors**

- Clearly excludes MTurk

Generalizability

- "The Generalizability of Survey Experiments" (Mullinix et al., 2015):
  - General point: survey experiments generalize...
  - "some convenience samples would be inappropriate such as a student sample where a moderator is age"

- **Any sample with a hard digital literacy cutoff is inappropriate for making generalizations about online behaviors**

- Clearly excludes MTurk

- Excludes even our over-sample of digital naives due to limitations of survey instruments we never thought existed

Results about Generalizability are not Generalizable

- What should be the default expectation for treatment effects?

Results about Generalizability are not Generalizable

- What should be the default expectation for treatment effects?
  - "Heterogeneous until proven/theorized to be homogeneous"?

Results about Generalizability are not Generalizable

- What should be the default expectation for treatment effects?
  - "Heterogeneous until proven/theorized to be homogeneous"?
  - Most extant heterogeneous effects are post-hoc and underpowered

Results about Generalizability are not Generalizable

- What should be the default expectation for treatment effects?
  - "Heterogeneous until proven/theorized to be homogeneous"?
  - Most extant heterogeneous effects are post-hoc and underpowered
- WEIRD subjects in psych—largely disproven by Many Labs 2

Results about Generalizability are not Generalizable

- What should be the default expectation for treatment effects?
  - "Heterogeneous until proven/theorized to be homogeneous"?
  - Most extant heterogeneous effects are post-hoc and underpowered
- WEIRD subjects in psych—largely disproven by Many Labs 2
- **HUGE** heterogeneity $\equiv$ low generalizability for much of development econ

Results about Generalizability are not Generalizable

- What should be the default expectation for treatment effects?
  - "Heterogeneous until proven/theorized to be homogeneous"?
  - Most extant heterogeneous effects are post-hoc and underpowered
- WEIRD subjects in psych—largely disproven by Many Labs 2
- **HUGE** heterogeneity $\equiv$ low generalizability for much of development econ
- My (potentially controversial) view: effect homogeneity in lab/survey experiments is better evidence for their artificiality than evidence that "treatment effects are homogeneous"

Results about Generalizability are not Generalizable

- What should be the default expectation for treatment effects?
  - "Heterogeneous until proven/theorized to be homogeneous"?
  - Most extant heterogeneous effects are post-hoc and underpowered
- WEIRD subjects in psych—largely disproven by Many Labs 2
- **HUGE** heterogeneity $\equiv$ low generalizability for much of development econ
- My (potentially controversial) view: effect homogeneity in lab/survey experiments is better evidence for their artificiality than evidence that "treatment effects are homogeneous"
- Internet-based surveys that study online behavior are thus "more field-y" than internet-based surveys that study something else
  - "Users over 65 shared nearly 7 times as many articles from fake news domains as the youngest age group" (Guess, Nagler, and Tucker, 2018)

Results about Generalizability are not Generalizable

- What should be the default expectation for treatment effects?
  - "Heterogeneous until proven/theorized to be homogeneous"?
  - Most extant heterogeneous effects are post-hoc and underpowered
- WEIRD subjects in psych—largely disproven by Many Labs 2
- **HUGE** heterogeneity $\equiv$ low generalizability for much of development econ
- My (potentially controversial) view: effect homogeneity in lab/survey experiments is better evidence for their artificiality than evidence that "treatment effects are homogeneous"
- Internet-based surveys that study online behavior are thus "more field-y" than internet-based surveys that study something else
  - "Users over 65 shared nearly 7 times as many articles from fake news domains as the youngest age group" (Guess, Nagler, and Tucker, 2018)
  - Evidence from a 61-million person experiment: "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" (Bond et al., 2017)

Internet-Based Surveys for Online Behavior

- Research design needs to incorporate digital literacy

Internet-Based Surveys for Online Behavior

- Research design needs to incorporate digital literacy
- Should always run a pilot on low-digital literacy populations

Internet-Based Surveys for Online Behavior

- Research design needs to incorporate digital literacy
- Should always run a pilot on low-digital literacy populations
  - Be sure to collect data on devices eg mobile / desktop / tablet (Searles et al., 2017; Searles and Dunaway, 2017)

Internet-Based Surveys for Online Behavior

- Research design needs to incorporate digital literacy
- Should always run a pilot on low-digital literacy populations
  - Be sure to collect data on devices eg mobile / desktop / tablet (Searles et al., 2017; Searles and Dunaway, 2017)
  - Is that meant to be part of the experiment? Does it covary with other demographics?

Internet-Based Surveys for Online Behavior

- Research design needs to incorporate digital literacy
- Should always run a pilot on low-digital literacy populations
  - Be sure to collect data on devices eg mobile / desktop / tablet (Searles et al., 2017; Searles and Dunaway, 2017)
  - Is that meant to be part of the experiment? Does it covary with other demographics?
- Current case: could easily have made the stories appear in-line

Internet-Based Surveys for Online Behavior

- Research design needs to incorporate digital literacy
- Should always run a pilot on low-digital literacy populations
  - Be sure to collect data on devices eg mobile / desktop / tablet (Searles et al., 2017; Searles and Dunaway, 2017)
  - Is that meant to be part of the experiment? Does it covary with other demographics?
- Current case: could easily have made the stories appear in-line
- Don't assume that everyone who misses an attention check is producing garbage data

Internet-Based Surveys for Online Behavior

- Research design needs to incorporate digital literacy
- Should always run a pilot on low-digital literacy populations
  - Be sure to collect data on devices eg mobile / desktop / tablet (Searles et al., 2017; Searles and Dunaway, 2017)
  - Is that meant to be part of the experiment? Does it covary with other demographics?
- Current case: could easily have made the stories appear in-line
- Don't assume that everyone who misses an attention check is producing garbage data
- Include at least one open-response question

Internet-Based Surveys for Online Behavior

- Research design needs to incorporate digital literacy
- Should always run a pilot on low-digital literacy populations
  - Be sure to collect data on devices eg mobile / desktop / tablet (Searles et al., 2017; Searles and Dunaway, 2017)
  - Is that meant to be part of the experiment? Does it covary with other demographics?
- Current case: could easily have made the stories appear in-line
- Don't assume that everyone who misses an attention check is producing garbage data
- Include at least one open-response question
- Better subject experience for everyone = better data

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.

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.".

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* .

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.

Searles, Kathleen, and Johanna Dunaway. 2017. "News and Information Loss in the Mobile Era." *Working Paper* .

Searles, Kathleen, Mingxiao Sui, Paul Newly, and Johanna Dunaway. 2017. The Limits of Digital Citizenship: Constraints on News Consumption and Recall in the Mobile Setting. In *Unpublished Manuscript*.

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