Digital Literacy and Online Political Behavior*

Andy Guess[†] and Kevin Munger[‡]

April 7, 2020

Abstract

Digital literacy is receiving increased scholarly attention as a potential explanatory factor in the spread of online misinformation. As a concept, however, it remains surprisingly elusive, with little consensus on definitions or measures. Building on work in communication studies and sociology, we provide a unified framework of digital literacy for political scientists and introduce survey items to measure it. Using a novel purposive sampling approach, we then validate our measure against real-world benchmarks of ground truth. There exists substantial variation in levels of digital literacy in the population, which we also document is correlated with age and could confound observed relationships. However, this is obscured by researchers' reliance on online convenience samples that select for people with computer and internet skills. We discuss the implications of this sample selection bias for effect heterogeneity in studies of online media effects on political behavior.

Click here for latest version.

*This research was supported by a Facebook Integrity Foundational Research Award and approved by the Princeton University Institutional Review Board (Protocol 11745). The authors are grateful to Eszter Hargittai for discussions and feedback on this project as it developed. We are indebted to Melissa Morrone at the Brooklyn Public Library and Kelsey Ockert and Morgan Taylor at the Princeton Public Library for their advice and help recruiting respondents. Thanks to Drew Dimmery for providing comments. Special thanks to Nejla Asimovic and Jacob Sweetow for outstanding research assistance.

[†]Assistant Professor, Department of Politics and Woodrow Wilson School, Princeton University. Email: `aguess@princeton.edu`

[‡]Assistant Professor, Department of Political Science and Social Data Analytics, Penn State University. Email: `kevinmunger@gmail.com`

1 Introduction

The rush to study online misinformation in the wake of the 2016 U.S. presidential election produced an unusual degree of research in parallel. Multiple research teams tackled the issue, using distinct data sets and statistical models. They reached two main conclusions. First, the overall consumption of low-quality, untrustworthy, or "fake" news was small. Second, the consumption and sharing of this fake news was very unequally distributed — and much higher among older internet users.

Using survey data linked with Facebook data, Guess, Nagler, and Tucker (2019) report that "users over 65 shared nearly 7 times as many articles from fake news domains as the youngest age group" during the 2016 campaign. Barbera (2018) finds that people over 65 shared roughly 4.5 as many fake news stories on Twitter as people 18 to 24. Grinberg et al. (2019) find that people over 65 were exposed to between 1.5 and 3 times as much fake news on Twitter as the youngest people; the slope depends heavily on partisanship. And matching Twitter users to voter files, Osmundsen et al. (2020) find that the oldest age group was 13 times more likely to share fake news than the youngest.

Misinformation is not the only realm of online political behavior where age-based heterogeneities of this magnitude have been recorded. The most prominent digital voting experiment, conducted in concert with Facebook in 2010, finds similar results (Bond et al. 2012, 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."

The magnitude and consistency of age-based heterogeneities in online political behavior suggests that there are some qualitative differences in how people experience digital media effects. "Age," however, is not a theoretically interesting covariate — despite considerable effort, it remains difficult to intervene on, and it is hard to conceive how age *per se* could have an impact on how people engage with digital media.[1] It is thus likely that age is confounded with another individual-level trait not commonly measured in studies of political behavior.

Motivated by this possibility, we explore the theoretical construct of *digital literacy* and its relationship to "computer skill" and "internet skill." We discuss the history of these terms and the differences between them and argue for the adoption of digital literacy as a concept of general interest to social scientists, which we define as online information discernment combined with the basic digital skills necessary to attain it.

[1] It is possible that there are physical elements related to aging that do play a role. Old age is associated with declines in eyesight and fine motor skills, which could plausibly be related to the capacity to quickly observe source cues or conduct independent verification in new browser tabs. These problems are likely much more acute for mobile web users (Dunaway et al. 2018).

There is significant anecdotal evidence that these skills covary with age (at least in the United States) and that older people are less digitally literate. We agree with previous research that finds significant variation in digital literacy at every age; Hargittai and Micheli (2019) argues that “contrary to claims about so-called digital natives, many youth lack information-evaluation abilities” (see also Metzger et al. 2015). While this naïve dichotomy of “digital natives” and “digital immigrants” is false, we do find evidence of a strong correlation between age and digital literacy in both a large national sample and an opt-in sample recruited via Facebook ads. Crucially, however, we find no correlation between age and digital literacy in a sample recruited from Amazon’s Mechanical Turk (MTurk) platform.

We perform a “horse race” between survey scales that have been proposed to measure digital literacy in order to test which is best at discriminating between respondents from samples recruited from groups that we identify as being ex ante high (employees of tech companies) or low (participants in introductory computer skills classes) on this characteristic. A well-known scale introduced by Hargittai (2005) performs best overall, although individual items from both another existing scale (Sundar and Marathe 2010) and one that we originate show promise. Notably, paradata measures such as *browser version* discriminate powerfully between our samples, although it is unclear whether this finding can translate to a reliable proxy for digital literacy.

Our findings suggest that existing surveys utilizing the Hargittai battery meaningfully capture digital literacy as we conceptualize it. At the same time, we recommend that researchers incorporate items from the other scales where possible in order to ensure that measures adequately separate those at the very low and high ends of the scale. To guide practice, we suggest a selection of 10 items grouped into two survey grids, although these can be tailored to specific research purposes or refined with future validation efforts building on what we demonstrate here. Returning to one of the motivations for our investigation, we document with a large national sample that digital literacy (as measured by a shortened version of the Hargittai battery) does indeed covary with age, with older Americans having lower levels on average.

We conclude with a discussion of implications for the use of online convenience samples in studies of online political behavior, and how the concept of digital literacy can usefully contribute to several other literatures likely impacted by recent shifts in the online information environment.

2 Skill and Media Modality: A Brief History

The rapid adoption of the internet and social media in the United States has changed the media landscape. But by most ways of counting, the spread of first radio and then television through society happened more rapidly, and each had a larger impact on the daily media diets of the average consumer (Prior 2007).

This was only possible because radio and television are *easy* to consume: everyone is trained to listen to human speech from birth, or to listen to speech paired with a visual stimulus. Other than language barriers or the rare niche program, anyone could turn on their radio set or television and share an experience with anyone else listening.

This audience-capacity line of thinking is inspired by Bennett and Iyengar (2008)'s magisterial retrospective on the history of communication scholarship, which argues that the analytic perspectives that scholars find compelling at a given time are driven by their technological and social framework. They argue, for example, that communications theorists of the modernity of the early 20th century were grappling with the novelty of mass society made possible by these broadcast technologies. Analyzing novel, pressing developments is a great angle for scholarship, but this first wave may have been *too* influential — the modernist paradigm continued to dominate theoretical inquiry long after the premises that motivated it failed to obtain.

The internet is different in every respect. It took many years to develop and has been constantly mutating. During this time, the only consumers tended to also be producers. Only people with specialized skills could even access online media, which was at first only written text. This early online media was thus produced and consumed by a small number of professionals in academia and tech companies, gradually expanding to geeky hobbyists with enough free time and disposable income to purchase and use the necessary, complicated hardware (Abbate 2000; Burke et al. 2005).

For the past 50 years, television has been the dominant modality for political media. The total supply of television content is constrained by the cost of production and distribution. And televisual content requires minimal skill to appreciate. So, with a small number of notable exceptions[2], the dominant theories of media effects were focused on *homogeneity.*

The internet inverts this. Central tendencies are simply not that informative. Individuals' experiences are so distinct that heterogeneity should be the baseline expectation.

The technological affordances of online media allow for a much greater variety of content,

[2]In addition to the widespread theory that media effects are heterogeneous in partisanship, Prior (2007) and Arceneaux and Johnson (2013) establish the importance of audience preferences for entertainment limiting the total reach of political media. Mutz (2015) makes a similar case for studying the heterogeneity in audiences' conflict avoidance in understanding the reach of uncivil cable talkshows.

expanding the scope of "politics" beyond the evening news or cable talk show. Furthermore, this variety has increased over time, as more different types of people and organizations produce that content. But this is only one source of increased heterogeneity of media effects due to the expanded diversity of the audience for online media. Unlike a radio or television broadcast, where the range of the experiences among adults exposed to a given piece of content is limited, the range of the experiences among adults exposed to online media is *extremely wide*, at least for the internet audience of the late 2010s.

These *experiences* are created at the intersection of a media consumer and a piece of online media. The classic model of political sophistication in Luskin (1990) conceives of the acquisition of political information as a process with three inputs: access to that information, the motivation to acquire it, and the ability to process it.

Access has clearly increased with the web, and Sood and Lelkes (2018) claim that rising education levels suggest increased ability, at least in the aggregate. The key complication, we argue, is that the *difficulty* of acquiring (true) political information on social media is much higher than in other modalities—and that *ability* is currently highly heterogeneous.

Our aim is to translate a large literature on this internet *ability* from other disciplines into the argot of political science. This literature begins with Hargittai (2001), who describes the "Second-Level Digital Divide." To that point, concern about the "Digital Divide" was over differential access to the internet. Hargittai's method was participant observation: she sat behind people and watched them use the internet to search for information.

This simple act was revelatory (and is a tool that could still be fruitfully used by researchers to study novel internet technologies). People used the internet *very differently.* They varied considerably in their capacity to find factual information, and in their speed in doing so. Hargittai conceptualized this as "online skills"—a concept that refers explicitly to the actual, validated ability to " to efficiently and effectively find information on the Web." There are of course a variety of different specific skills that comprise one's overall internet ability, and other skills involving related technologies (e.g. "computer skills," related to hardware).

One drawback of this measure of "internet skills" is that directly observing these skills at scale is difficult. Instead, Hargittai (2005) develops and validates a survey measure of "digital literacy." This exercise demonstrated that a battery of survey questions that ask respondents to rate their familiarity with a series of internet terms was most predictive of the underlying internet skills of interest. "Digital literacy," in Hargittai's conception, refers to these survey questions, which are a proxy for internet skills.

3 Digital Literacy: Theoretical Building Blocks

The term *digital literacy* (or *digital media literacy*) is frequently invoked but rarely defined consistently.[3] Possibly for that reason, its boundary is porous: discussions of digital literacy tend to overlap with related concepts such as internet skill (Hargittai 2005), media literacy (Vraga and Tully 2015; Vraga et al. 2015), and digital inequality (DiMaggio, Hargittai et al. 2001). Unsurprisingly, then, research on the topic is dispersed across multiple disciplines ranging from sociology and communication (e.g., Koltay 2011; Kahne, Lee, and Feezell 2012) to library and information sciences (e.g., Robinson et al. 2000; Eshet 2004). Building on existing work, a contribution of this paper is to theoretically situate the concept of digital literacy and specify it in a way that translates to straightforward, valid measures suitable for use in political science.

We begin by postulating that being digitally literate means being able to reliably assess the credibility of information encountered online (e.g., Flanagin and Metzger 2007). This in turn depends on the ability to verify claims and look up answers to questions using a variety of strategies.[4] While this may seem similar to traditional notions of information fluency (Sharkey 2013), it is situated within people's digital environments and the social context they experience online. In other words, we do not conceive of digital literacy as something specifically related to formal information sources (news organizations) or topics of continued interest to scholars (politics, health). A key feature of multifaceted environments is the constant need to assess the credibility of claims and requests, both formal *and* informal — not only while perusing social feeds and news headlines, but also in everyday encounters like phishing attempts and spam in one's email inbox. Whether routine or connected to specific informational needs, all of these tasks require judging what's on the screen in front of them: separating the trustworthy (genuine email communication, professional-quality news, credible claims by friends and acquaintances) from everything else.

Becoming proficient at these judgments requires experience as well as the fundamental skills needed to enable it. In the realm of routine tasks, recognizing a phishing attempt requires familiarity with the mechanics of email. Checking the credibility of a source might require opening a new browser tab (on a desktop or laptop computer) or switching apps (on a mobile or tablet device) — building blocks so essential that they are often bundled alongside higher-level competencies when taught in libraries and schools.

[3] "The indistinct use of the term causes ambiguity, and leads to misunderstandings, misconceptions, and poor communication among researchers" (Eshet 2004); Bawden (2008) describes it more bluntly as "a topic whose terminology is very confused" (17).

[4] Such strategies for critically evaluating online information include "lateral reading" (Wineburg and McGrew 2018) and the "SIFT" technique (`https://hapgood.us/2019/06/19/sift-the-four-moves/`).

At its core, then, digital literacy consists of a *skills* component and an *information literacy* component. This is implied in some existing scholarship. In a foundational book on the topic, Gilster (1997) asserts that "digital literacy is about mastering ideas, not keystrokes" (15), suggesting the primacy of a more general kind of information literacy. But in the same book, Gilster defines core competencies of digital literacy which include the ability to search the internet and navigate hypertext. Taking the link between skills and information literacy to its logical conclusion, one could argue that literacy is itself a kind of skill. Hargittai and Micheli (2019) identify 10 dimensions of "internet skills," such as interpersonal communication and managing privacy, which collectively amount to a digital citizenship test for the 21st century. One of Hargittai and Micheli's skills dimensions is "ability to find and evaluate information," which corresponds most directly to notions of information literacy or discernment.

This twin conception of digital literacy — information discernment combined with the basic digital skills necessary to attain it — has two advantages. It unifies disparate strands of scholarship while establishing boundaries that separate it from related ideas such as digital savvy and news literacy. At the same time, it allows for tractable measurement: Instead of attempting to detect information literacy, we focus on the skills component as a useful proxy for the whole.

The question of how to define and measure digital literacy returns us again to political sophistication (Luskin 1990), which similarly bundles several components (i.e., knowledge, interest, and abstract reasoning ability). Reviewing the many existing approaches to measuring the concept, Luskin (1987) proposed a type of "recognition and understanding" measure (Converse 1964) that grades respondents on a series of factual questions about the ideological leanings of the parties. This measure intentionally taps the knowledge element of sophistication ("what") rather than assessing directly people's belief systems ("how") or the extent to which they think about politics at all ("how much"). As Luskin argues, "Empirically ... there should be a strong relationship between the 'how much' and 'how' and some factual aspects of the 'what"' (881) — in other words, while it may be theoretically possible to have a sophisticated political belief system built on basic factual misconceptions, it is difficult to envision in practice. Likewise, someone who lacks the digital acumen to download and send a file may be well versed at searching for reliable information online, but the combination would be unusual.[5]

As political sophistication came to be an important moderator in studies of retrospective

[5] An interesting corollary is that the skills component lends itself to direct behavioral observation without the need for self-reported survey responses. However, as we discuss below, translating that to usable measures is a nontrivial challenge.

voting and preference formation, we argue that digital literacy is a crucial factor in online political behavior whose role has to date been obscured by disciplinary practices designed for an earlier media-technological environment. But how should digital literacy — in particular, the skills component we take as a tractable proxy for the whole — be measured?

4 Validation Through Purposive Sampling

A person's "ground truth" level of computer skill can be measured explicitly through participant observation, the method used by Hargittai. However, there are a wide variety of skills which are potentially relevant to political scientists, and this measurement approach is difficult to scale across many populations. We thus adopt a benchmarking strategy in which we compare survey responses drawn from populations identified ex ante as likely to possess target levels of digital literacy. This is analogous to validation studies that benchmark survey measures of digital media and social media use against respondents' digital trace data (Guess 2015; Munger et al. 2018), although our approach here is more indirect.

The ideal survey measure has several properties. First, it captures variation at both high and low levels of skill. Second, it reflects the process that causes the effects of internet and digital media use to vary between individuals. And third, it is high in "temporal validity": the relationship between the survey-based proxy and the underlying skills of interest is constant across time (Munger 2019).

We preform a "horse race" to see which of the several proposed survey-based proxies is best according to these three criteria. Our validation takes a novel form. We identify two distinct groups who we assume represent the poles of computer skill.

Our "high-skill" sample ($N = 83$) was constructed by asking friends and colleagues who work at technology companies to distribute the survey among their workforce. Navigating computer technology is central to these people's careers, and our assumption is thus that they are highly skilled.

In contrast, our "low-skill" sample ($N = 18$) is comprised of people who were taking introductory computer skills classes at the Princeton Public Library and the Brooklyn Public Library. One of us agreed to offer a course on R (to a more advanced audience) in exchange for the cooperation of some of these classes. Our assumption here is that people who were taking the time to learn the basics of computing are lacking in those skills.

The "horse race" competitors are three existing scales that have been theorized to be relevant to online media effects. The first is Hargittai's 21-question battery asking respondents to self-report their familiarity with a series of computer- and internet-related terms, on a 5-point scale (Hargittai 2009).

The second is the “power user” scale developed by Sundar and Marathe (2010). This scale consists of 12 questions asking respondents about how they interact with technology, on a 9-point agree-disagree scale. Examples include “I make good use of most of the features available in any technological device” and “Using information technology makes it easier to do my work.” The lack of proper nouns in this scale is likely to make it higher in “temporal validity,” although this is merely an assumption that needs to be tested.

To these two, we added a third scale of out own construction: the “low end” scale. This scale adopted the intuition behind the “power user” scale, which was designed to separate high-skill users from everyone else. Our scale, in contrast, was designed to separate low-skill users from everyone else. The other scales can be found in the Appendix, but we report our novel “low end” scale here:

- A lot of the things I see online confuse me.
- I have problems with viruses and malware on my computer.
- I have trouble finding things that I’ve saved on my computer.
- I rely on family members to introduce me to new technology.
- I have professionals (such as the Geek Squad) or family members take a look at my computer when something isn’t working.

In addition to these scales, we captured a series of technical details recorded by Qualtrics, including the respondents’ web browser, browser version, operating system, and screen resolution. These are all choices that web users make, even if some don’t realize that they have a choice. Our theory was that people who use default web browsers and who do not keep them up to date are likely to be less skilled.

To conduct the horse race, we used a random forest classifier to predict whether each respondent was drawn from the high or low skill sample. We used the default values for the tuning parameters and ran each model with 1,000,000 trees. As a result, the uncertainty from the simulations is reduced to several orders of magnitude smaller than the estimates in all of the analyses reported below.

Figure 1 displays the results of this analysis with the aggregate measures, ranking them in order of the increased performance. The left panel displays the measures in terms of their contribution to classification accuracy, while the right panel plots them by the decrease in the average node impurity in the final leaf nodes. The results are broadly similar across the two metrics, with the exception of the importance of the 12-question power user scale, which plays a much larger role in increasing node purity.

Hargittai's 21-question digital literacy scale provides the most information to the classifier, followed by the browser version recorded by Qualtrics. The magnitude of the latter is particularly noteworthy because this is only one variable's worth of information. Recording the screen resolution (which in practice is information about whether the user is on a computer or a mobile device, and if the latter, the make and model of that device) also improves classifier performance, as does the user's operating system. Merely recording which browser is used provides little improved performance.

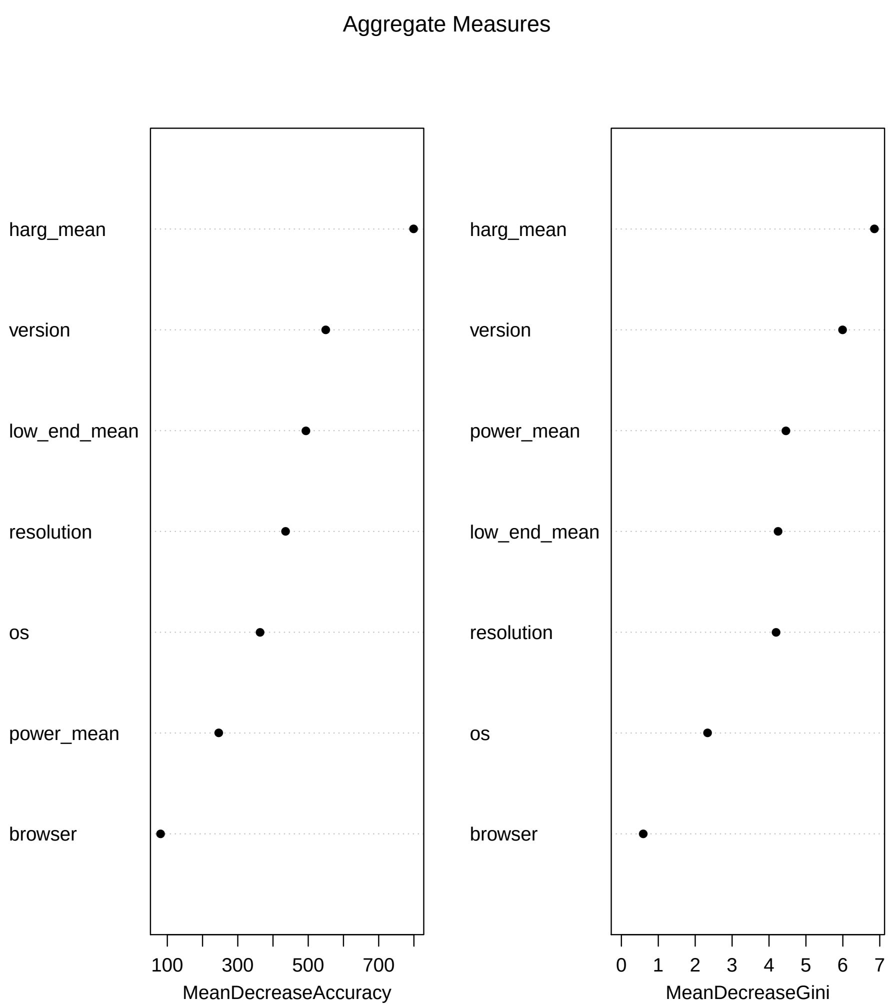
Because random forest is flexible and only uses a small number of predictors in each iteration, we can re-run the analysis with each of the aggregate measures broken down into its constituent survey questions. Figure 2 displays this analysis, this time without the Qualtrics metadata, each of which significantly outperforms any one survey question.

Our disaggregated analysis confirms the previous findings that the Hargittai scale performs best overall: six of the 10 most predictive survey items come from that battery. Three come from the Power User scale, and our Low End scale contributes one item.[6] The highest-performing item by far — recognition of the term "RSS" — is notable because its importance to internet fluency is relatively time-invariant. RSS stands for "Really Simple Syndication" and provides a standard way for websites to publish machine-readable content updates using a simple markup language. This was valuable for early internet enthusiasts and news junkies who could use feed readers (such as Google Reader) or custom-built code to sift through information from a large number of sources. Even in today's tech-giant-dominated, social-media-saturated era, RSS feed icons can be found on most publishers' websites and the standard's use continues among the dedicated. This post-hoc analysis suggests that our finalized measure will contain items with high temporal validity, a key criterion for social science measures related to internet behavior.

Similar arguments can be made for "JPG," which references a commonly used image format introduced in the early 1990s. As with RSS, recognition of JPG likely has high temporal validity, although it is a lower-difficulty item. Finally, it is interesting that one of the most predictive Power User items states that "Using information technology makes it easier to do my work," given that respondents who said that this often applies to them most likely worked in technology firms. We are agnostic as to whether this is a quirk of our sample.

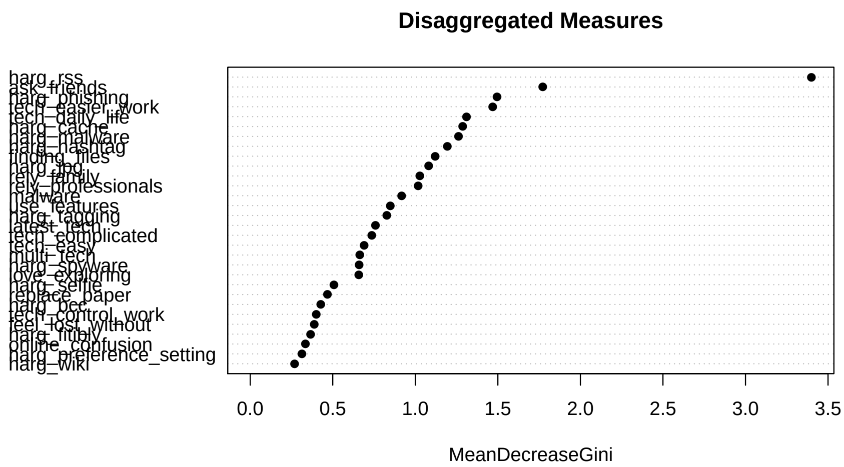
[6] We view this as suggestive evidence that with more respondents from low-digital-literacy populations, the random forest model would pick out additional items capable of discriminating at the lower end.

Figure 1: Increased Sample Classification Accuracy, Aggregated Measures



Results of a Random Forest classification of 101 respondents as either from the High DL Sample ($N = 83$) or Low DL Sample ($N = 18$). Each row in the left panel represents the decreased classification accuracy when that measure is removed from the analysis, while each row in the right panel represents the decrease in the average purity of leaf nodes.

Figure 2: Increased Sample Classification Accuracy, Disaggregated Measures



Results of a Random Forest classification of 101 respondents as either from the High DL Sample (83 respondents) or Low DL Sample (18 respondents). Each row represents the decrease in the average purity of leaf nodes when that measure is removed from the analysis.

5 Within-Sample Validation

As an additional validation of the measure, we included in our survey instrument a series of three information retrieval tasks. These are online information-related tasks of interest to political scientists; to our knowledge, this is the first internet-based survey implementation of these types of tasks.

Internet-based surveys can be a drawback for scholars studying political knowledge because they enable the possibility that respondents will "cheat" and look up the answers online (Clifford and Jerit 2016; Smith, Clifford, and Jerit 2020). We use this feature to our advantage, explicitly instructing respondents to look up the answers and intentionally asking questions which are sufficiently obscure to U.S.-based respondents that almost no one could know the answer already.

The three information retrieval questions were:

- Who is the Prime Minister of Croatia? (Andrej Plenković; the presence of the concluding diacritical means they copy+pasted the answer)
- What is the capital city of Malawi? (Lilongwe)
- What is the only US National Park that begins with the letter "T"? (Theodore Roosevelt National Park)

To supplement the small purposive samples, we replicated the survey on two larger samples drawn from online sources commonly used in political science research.

One is a sample recruited from MTurk ($N = 503$), a now-ubiquitous source of subjects for many types of social science experiments. The MTurk interface is somewhat difficult to navigate, and qualitative evidence from Brewer, Morris, and Piper (2016) suggests that this barrier prevents low-skill people from using the platform. As a result, we expect this sample to skew to the high end of digital literacy and to contain a hard floor, below which there will be zero individuals.

The second follows Zhang et al. (2018) and Munger et al. (2020) and use a Facebook ad to recruit survey participants ($N = 451$). Facebook is generally very accessible, and many users have accounts created for them by younger, more tech-savvy relatives to make it easier to keep in touch. Redmiles, Chachra, and Waismeyer (2018) reports that certain groups of Facebook users are much more likely to click on spam, in particular, women and lower-skill users. Munger et al. (2020) report that similar groups are more likely to click on these Facebook recruitment ads. As a result, we expect this sample to skew toward the lower end of digital literacy.

The distributions of the correct answers to these three questions across these four initial samples comports with our expectations. On average, the tech-worker "high" sample does the best, and the computer-class "low" sample does the worst, as can be seen in Figure 3.

The first two questions are both similarly straightforward. A single Google search of the exact question will produce one of Google's auto-populated "answer boxes" with the correct answer. On these questions, our expectations of the relative performance was exactly correct: the MTurk sample did better than the Facebook sample, and they were between the high and low samples.

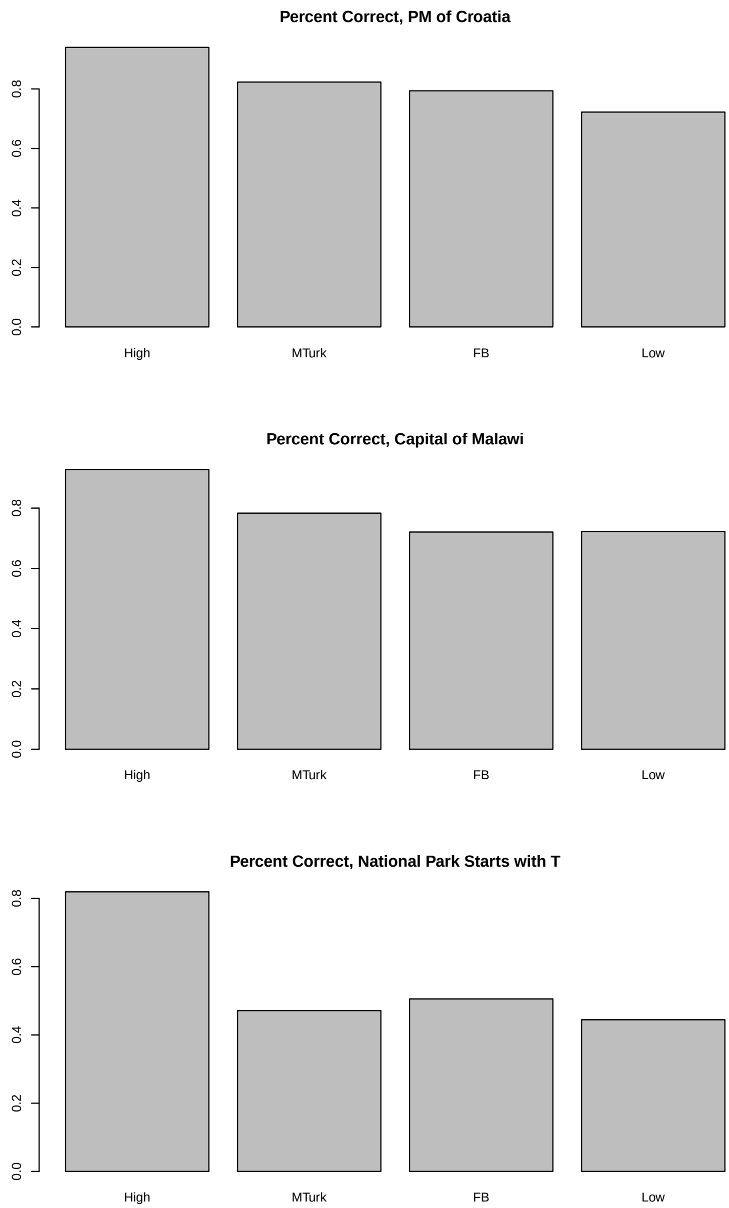
The third question was selected specifically to require slightly more skill at crafting a search term and navigating the options provided. Here, the only gap is between the "high" sample and everyone else. This suggests that lower-skill individuals, of which there should have been some across the other samples, had difficulty with this question.

We can also re-run the random forest analysis above to see whether the survey measures of digital literacy that best discriminated between the high and low samples are also good predictors of the capacity for information retrieval. Figure 4 uses the aggregate measures and the Qualtrics metadata. Here, the Hargittai digital literacy questions are consistently the most predictive of correctly answering the question. In each case, the power user scale is close behind. We also include an indicator for which sample the respondent was drawn from, "which_f" in the plots. Reassuringly, adding this variable does little to improve the prediction as to whether the respondent correctly retrieved the information, despite the wide variation in performance between the samples seen in Figure 3. This suggests that the other measures accurately capture the mechanisms underlying the baseline difference between samples. Figure 8 (Appendix) again disaggregates the scales into their constituent parts.

5.1 A Validated Measure of Digital Literacy

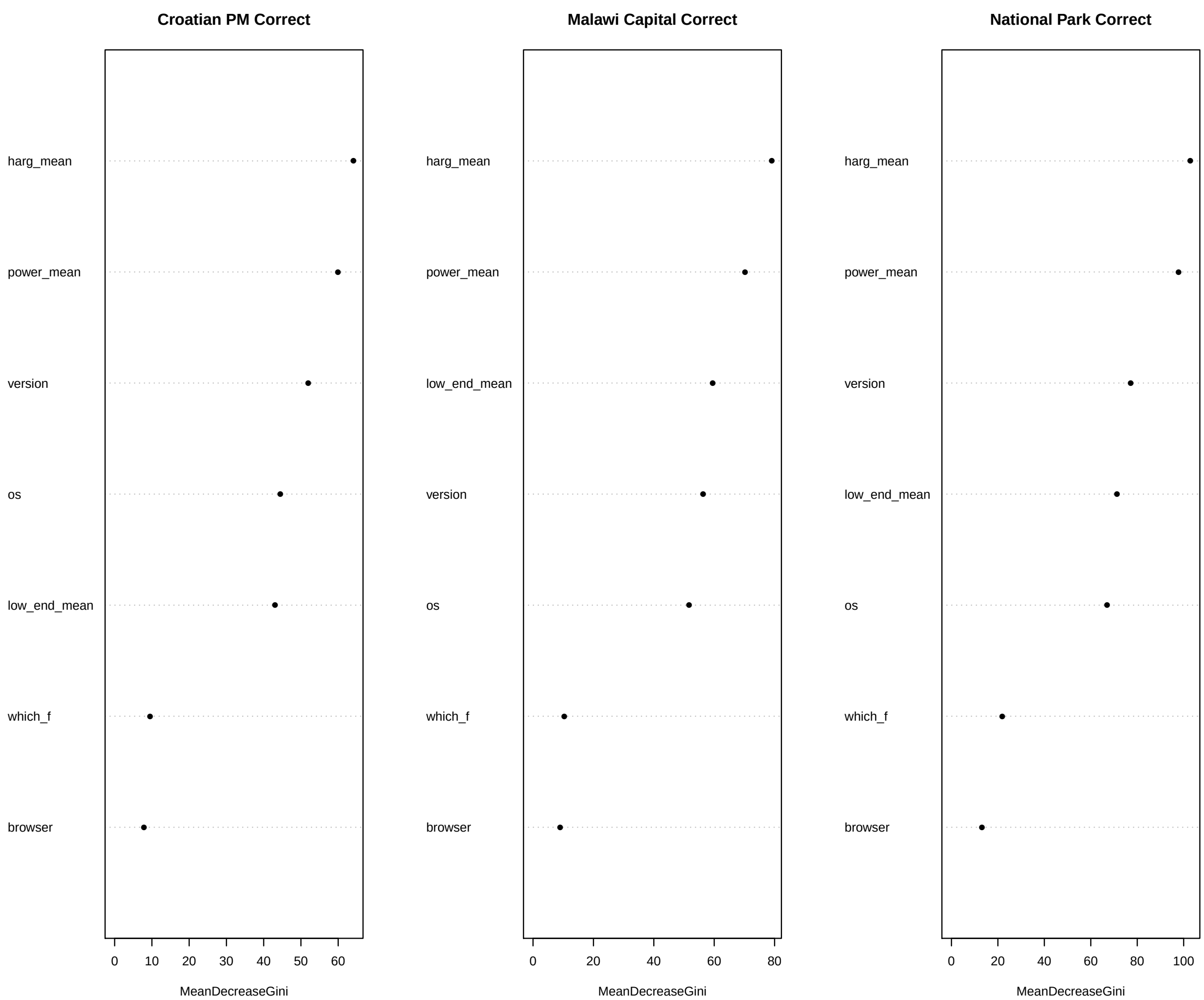
Using the most predictive items from our validation tests, we propose a new survey measure of digital literacy. It combines items from the Hargittai and Power User scales along with one item from our novel Low End battery adapted to a Likert scale, arranged in two survey grids:

Figure 3: Information Retrieval Accuracy Across the Four Samples



Percentage of respondents from each sample (Facebook sample, $N = 451$; High DL sample, $N = 83$; Low DL sample, $N = 18$; MTurk sample, $N = 503$) who correctly answer each of three information retrieval questions.

Figure 4: Increased Classification Accuracy of Information Retrieval, Aggregated Measures



Results of a Random Forest classification of 1070 respondents aggregated across all four samples. Each row represents the decrease in the average purity of leaf nodes when that measure is removed from the analysis. Each column takes the correct response of a different information retrieval question as the binary classification target of the Random Forest.

How familiar are you with the following computer and Internet-related items? Please choose a number between 1 and 5 where 1 represents "no understanding" and 5 represents "full understanding" of the item.

- Phishing
- Hashtag
- JPG
- Malware
- Cache
- RSS

Please indicate your agreement with the following statements on a scale of -4 = Strongly Disagree to 4 = Strongly Agree.

- I prefer to ask friends how to use any new technological gadget instead of trying to figure it out myself.
- I feel like information technology is a part of my daily life.
- Using information technology makes it easier to do my work.
- I often have trouble finding things that I've saved on my computer.

Responses can be additively converted to an overall scale. We especially recommend this set of 10 items to ensure adequate separation of respondents with very low or very high levels of digital literacy, although as our results have clearly shown, the Hargittai battery performs best overall. This suggests that existing surveys with that measure are still likely to have captured variation in digital literacy.

Note that we do not include in this scale the paradata recorded by Qualtrics about respondent hardware and software. Although some of these variables outperformed any of the individual survey questions, it is unclear how they could be reliably translated into a moderator of treatment effects. The browser version variable, for example, contains dozens of numeric codes that cannot easily be converted to a low-dimensional scale. Even worse, these hardware and software indicators are likely to change much more quickly than the underlying quantity of interest, so any relationship we document will decay quickly over time. While the fact that these data are "free" to unobtrusively collect is an appealing feature, making them tractable as usable measures is a topic for future research.

6 Distribution of Digital Literacy in Several Common Samples

Armed with our suite of measures, we now seek to categorize the distribution of digital literacy in three samples of interest to scholars of digital politics. The first two, drawn from Mechanical Turk and Facebook, were described above.

The third is a national sample recruited from Lucid ($N = 2,146$), an online survey platform recently attracting social scientists' attention (Coppock and McClellan 2019). We expect to find individuals in this sample that span the entire range of our scale, although it may still somewhat under-represent people at the lowest end of the digital literacy scale.

Figure 5 plots these distributions, along with either the low or high DL samples discussed above, depending on the measure.

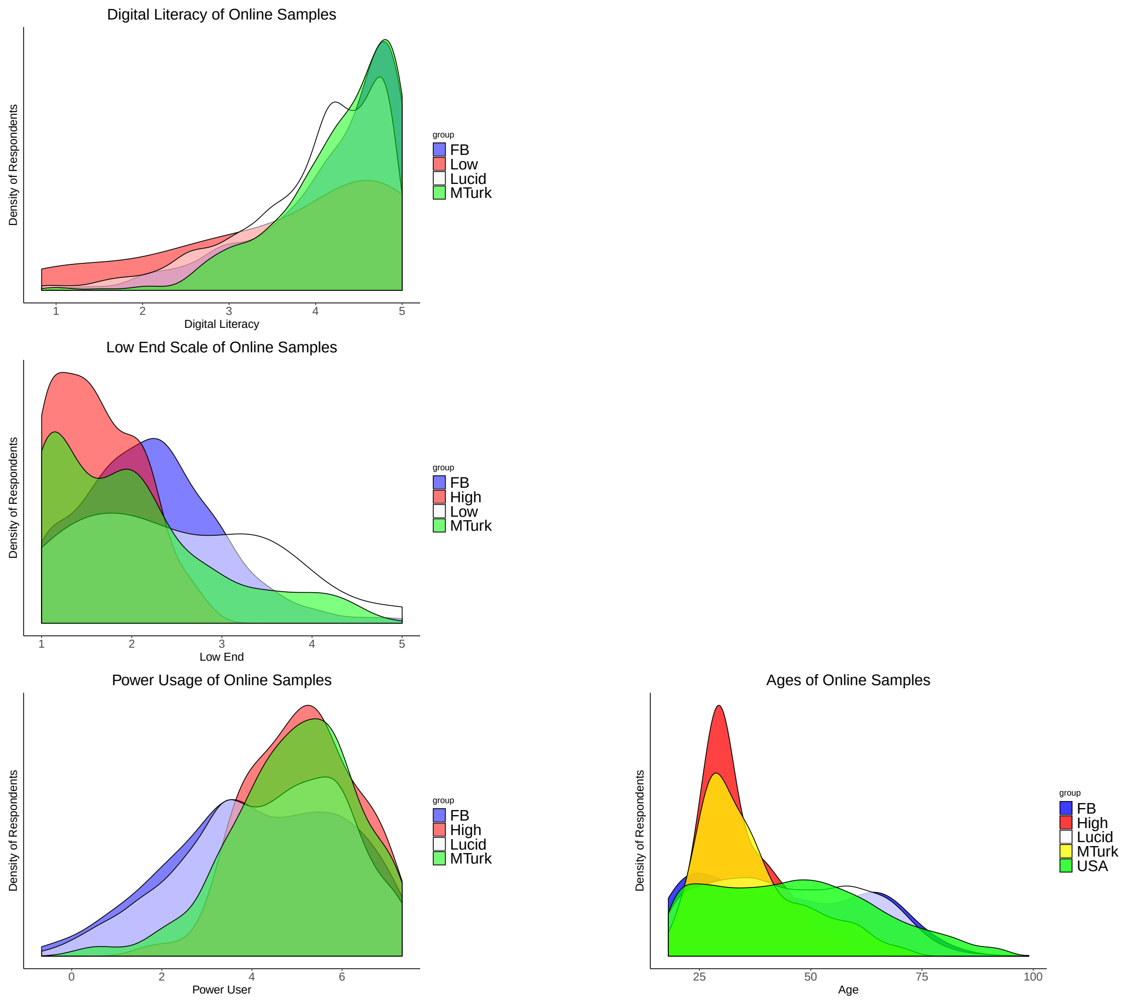
The first panel of Figure 5 analyzes the distribution of the demographic that first motivated this investigation into digital literacy: age. The results are stark. The age distributions for the MTurk and High DL samples are similar to each other and skewed far in the direction of the young, relative to the data from the 2010 US Census (in green, labeled "USA." The true age pyramid should of course be moved to the right to reflect the passage of 9 years). In contrast, the age distribution of the Facebook sample is actually slightly skewed to the elderly, with people in their 60s over-represented. The nationally-representative Lucid sample matches the census numbers closely, suggesting that this sampling was well-conducted.

The second panel of Figure 5 displays the results for the Hargittai digital literacy measure. The full 21-question scale was used for most of the samples, while the shortened 7-question scale was used for Lucid. As expected, the Low DL sample has a much flatter distribution, with some respondents who report not being familiar with nearly any of the internet terms. The MTurk sample is the least diverse in terms of digital literacy, with close to a hard floor at 2.5 on the 5-point scale. Only 1.4% of MTurk respondents were below this figure, compared to 16.7% of the Low DL sample, 5.1% of Facebook respondents, 5.8% of Lucid respondents, and 0 out of 83 respondents from the High DL sample.[7]

Panel 3 shows our novel Low End scale, designed to detect more variation among people

[7] The point of insufficient variation at the low end of digital literacy among MTurk workers was recently made by Hargittai and Shaw (2020), although not in the detail we provide here.

Figure 5: Distributions of Relevant Measures Across the Samples



Each plot represents the distribution of a given variable across 3 or 4 different samples (Facebook sample, $N = 451$; High DL sample, $N = 83$; Low DL sample, $N = 18$; MTurk sample, $N = 503$; Lucid sample, $N = 2,146$). The first plot is the distribution of "Digital Literacy" of respondents, as measured by the 21 Hargittai identification questions (shortened 7-question scale used for Lucid); the High DL sample is excluded because it is so far skewed to the right that the graph is unreadable. The second plot is the distribution of "Low End" of respondents, as measured by our novel 5-question scale, which we did not ask the Lucid respondents. The third plot is the distribution of "Power User" of respondents, as measured by the 12-question Power User scale; the Low DL sample is excluded because this measure is not designed to distinguish between people on the low end. The final plot is the distribution of ages of respondents; the Low DL sample is excluded because very few respondents answered the prompt with a number; "USA" refers to the population distribution from the 2010 Census.

lower in digital literacy. The modal High DL and MTurk respondents are at the bottom edge of this scale, although the support for the MTurk distribution does span the entire range. Facebook respondents are more normally distributed, while (as expected) the Low DL sample provides the highest proportion of respondents who score high on the Low End scale.

Finally, Panel 4 displays the results of the 12-question Power User scale. This is in many ways the most troubling result for Mechanical Turk: the distribution of this measure for our High DL sample of tech company employees is nearly identical to the distribution for MTurkers. In contrast, the measure is much more broadly distributed across the Facebook and Lucid samples, with a bump on the higher end for the latter.

The entirety of Figure 5 suggests a potentially serious methodological problem. The nationally representative studies of misinformation on social media during the 2016 election (Guess, Nagler, and Tucker 2019; Grinberg et al. 2019) demonstrated a correlation between age and sharing "fake news." We have suggested that this finding is due to the fact that age acts as a proxy for digital literacy, which is the key theoretical moderator.

We demonstrate this relationship in Figure 6, which also presents evidence for the crucial weakness of using MTurk samples for studying behavior moderated by digital literacy.

The top left panel of the first subfigure is evidence of the inverse relationship between age and digital literacy in the Facebook sample. The loess line of best fit demonstrates the non-linearity of this relationship; the line is flat (even sloping slightly upward at the low end) throughout the 18-50 age range, when it becomes significantly negative.

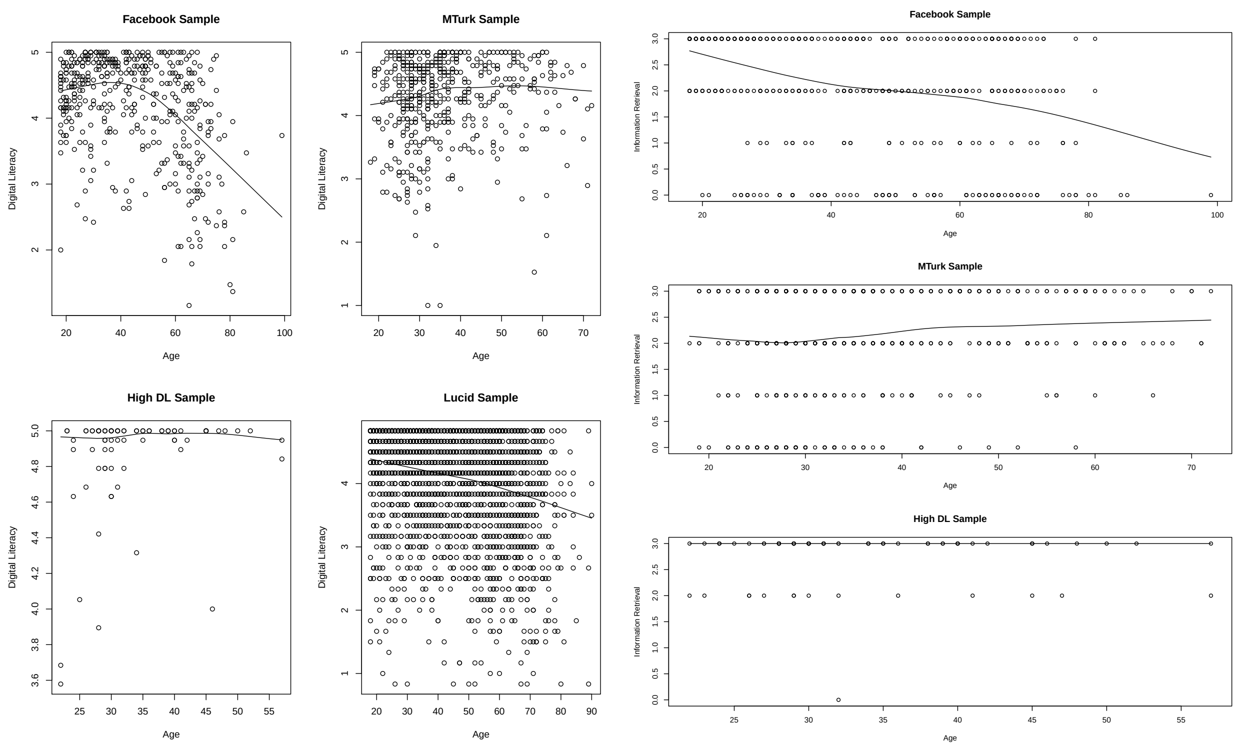
A similar relationship is found in the Lucid sample, portrayed in the bottom right corner of the first subfigure. In this sample, there is a negative, linear relationship between age and digital literacy throughout the sample. This represents the best evidence to date of a negative relationship between age and digital literacy in the United States today. The sample ($N = 2,146$) matches national distributions in age, gender, Hispanic, ethnicity, and region demographics.

However, the top right plot, using the Mechanical Turk sample, demonstrates *no correlation* between age and digital literacy. The average digital literacy of the 20-year-olds and 70-year-olds recruited via this platform is the same. A similar non-relationship is observed in the High DL sample in the bottom left plot, where (in accordance with the purposive sampling paradigm) a majority of respondents scored the maximum of the scale.[8]

Although we have demonstrated the validity of this scale for describing respondents'

[8] We cannot perform the same analysis on the Low DL scale because too few respondents provided us with their age. Only 7 provided a numeric response; several of the other respondents said things like (paraphrased for respondent privacy) "Too old for nosy questions" and "Old enough."

Figure 6: Relationship Between Age and Digital Literacy/Information Retrieval



Each scatter plot and loess represents the relationship between age and either Digital Literacy (the 21-question Hargittai scale (shortened 7-question scale used for Lucid); left column) or Information Retrieval (from 0 to 3, the number of successful internet searches for information; right column) in a given sample (Facebook sample, $N = 443$; High DL sample, $N = 83$; MTurk sample, $N = 503$; Lucid sample, $N = 2,146$). The Low DL sample is not present because too few respondents offered a numerical age.

level of internet skill for information retrieval, the second subfigure of Figure 6 presents as a robustness check the bivariate relationship between age and information retrieval in the various samples.

Again, there is a strongly negative relationship between age and information retrieval in the Facebook sample, although this time the slope is close to constant and negative across the age range. And again, there is no evidence of a negative relationship between age and information retrieval in the MTurk sample — in fact, there is a weak but *positive* relationship among people 30 and older.

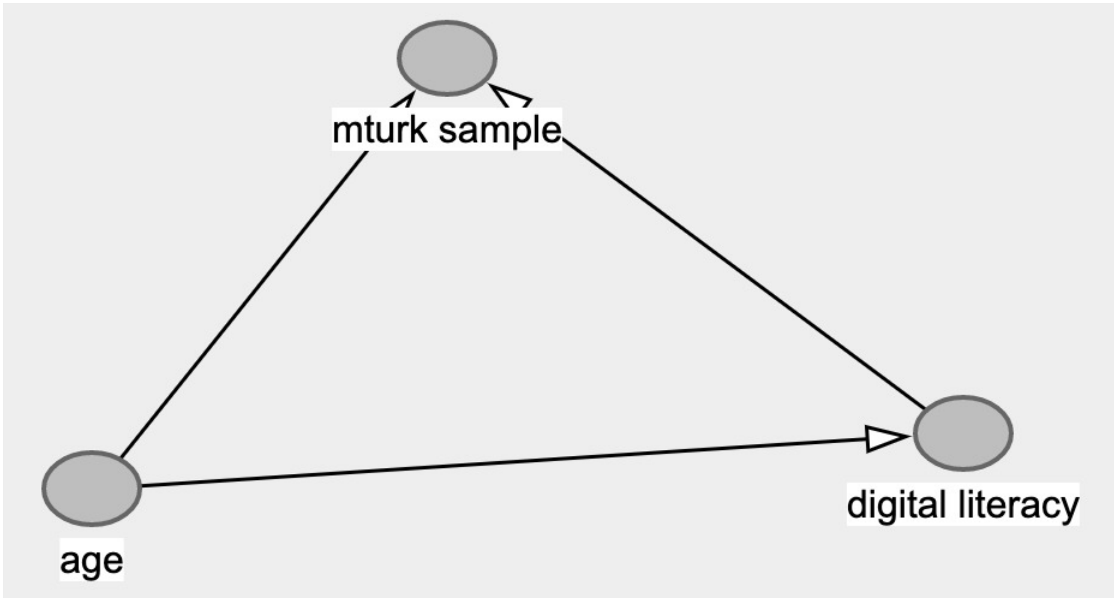
Presenting this another way, Table 1 displays a series of regressions with our index of three information retrieval questions as the dependent variable. Using the Facebook sample, the first two columns show the predicted negative relationship between age and information retrieval, even with the inclusion of Hargittai's aggregate measure. However, the next two columns replicate the analysis on the MTurk sample and find strong evidence of the *exact opposite* relationship: age is here positively correlated with information retrieval, with nearly identical magnitude and standard errors. In both samples, the measure of digital literacy is highly positively correlated with information retrieval.

Table 1: Comparing Age/Information Retrieval Relationship Across Samples

	Facebook Sample		MTurk Sample	
harg_mean		0.316^{***}		0.372^{***}
		(0.068)		(0.063)
age	-0.019^{***}	-0.014^{***}	0.014^{***}	0.012^{***}
	(0.003)	(0.003)	(0.004)	(0.004)
Observations	443	443	503	503
R^2	0.100	0.143	0.025	0.089
Adjusted R^2	0.098	0.139	0.023	0.086

This result can be described as an example of conditioning on a collider, a form of sample selection bias (Pearl 2009). The issue parallels the canonical examples of a non-relationship between SAT scores and GPA in selective colleges, or the non-relationship between height and performance among NBA players. In each case, individuals who score high on the former measure are more likely to be included in the sample; individuals who are included in the sample despite scoring low on the former measure are likely to possess other traits that contribute to their performance within the sample (see Knox, Lowe, and Mummolo

Figure 7: Graph of Selection Process



Forthcoming, for an illuminating discussion of this problem in the context of citizen-police interactions).

In the present case, age is predictive of both digital literacy and of being an MTurk worker, and digital literacy is also predictive of being an MTurk worker. Lower-digital-literacy internet users are less likely to be aware of the platform, and less likely to be able to navigate its non-intuitive interface.[9]

Figure 7 visualizes the problem using Pearl's graphical framework. This selection process means that the within the MTurk Sample, the causal relationship between age and digital literacy is broken.

As a result, experimental studies of digital media effects or online misinformation conducted on samples recruited from MTurk are unlikely to accurately estimate the role of age as a treatment effect moderator. Furthermore, as the first panel of Figure 5 demonstrates, there are close to zero MTurkers who are below a threshold of digital literacy (2.5, on the 1 to 5 scale). Given the highly unequal distribution of engagement and sharing of Fake

[9]Brewer, Morris, and Piper (2016) perform a qualitative analysis on a small sample of older adults who have not used Mturk, signing them up for the platform and interviewing them about the experience. Although these subjects generally reported being comfortable using computers, they could not complete basic tasks on Mturk:

> *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?"... after a new tab was opened. Also, participants often forgot the instructions immediately upon opening the new window.* (p.2251)

News during the 2016 election, concentrated among the elderly, it is likely that the most important respondents (who would be influential observations in any statistical analysis) are precisely those who are structurally excluded from MTurk.

7 Discussion: Literacies and Heterogeneity in Political Science

In this paper we argue that taking digital media seriously requires moving away from the assumption of media effect homogeneity developed in the broadcast era. Specifically, important heterogeneity in digital media effects can be captured through the digital literacy construct. We conceptualize digital literacy as having two intertwined components — information literacy and computer/web skill — and provide a validated survey measure based on using skill as a proxy for the whole.

We demonstrate that digital literacy varies considerably among populations frequently used by political scientists. In particular, we show that samples from Mechanical Turk contain vanishingly few low-digital-literacy respondents — precisely the population among whom we might expect to find the largest persuasive effects of digital media messages (including misinformation). We recommend that MTurk not be used for studies of digital media effects in the absence of strong theoretical reasons to expect effect homogeneity.

Other disciplines are grappling with the implications of substantively meaningful new forms of heterogeneity for common research practices. In psychology, the Cognitive Reflection Test (CRT) (Frederick 2005) was developed to measure "analytic cognitive style," a disposition that has been shown to covary with psychological and political attributes (e.g., less reflection is associated with paranormal belief and religious belief; see Pennycook et al. 2012, 2016). More recently, analytic cognitive styles have been shown to predict lower perceived accuracy of fake news headlines on social media (Pennycook and Rand 2019). Although CRT scores appear to be fairly stable over time (Stagnaro, Pennycook, and Rand 2018), there is evidence that the standard practice of relying on more experienced workers on MTurk could "conceivably inflate CRT scores observed in online samples and undermine the predictive accuracy of the CRT" (Chandler, Mueller, and Paolacci 2014, Study 2). This example similarly illustrates how endogenous selection bias can affect inferences: In this case, attentiveness or other forms of "non-naïveté" predict both inclusion in "high-quality" MTurk samples and high CRT, which is itself an important psychological moderator.

A broader question is how political science should handle novel online platforms for subject recruitment. The case of MTurk is illustrative. Berinsky, Huber, and Lenz (2012) "introduced" the platform to the discipline, and while their validation work is careful and nuanced, it has frequently been cited to imply that MTurk is generally "okay to use."[10] The article made a large impact and is overwhelmingly the most cited article published in *Political Analysis* in the past decade.

Numerous papers have argued about the merits of MTurk as a respondent pool for experimental studies. Coppock (2019), the most recent and comprehensive, establishes that a wide variety of experimental findings generated from nationally representative surveys can be replicated using MTurk samples. The reasoning is that since treatment effects are similar for many different types of subjects, the composition of online convenience samples along commonly measured demographic and political characteristics is relatively unimportant for recovering experimental treatment effects. Crucially, however, results about the generalizability of MTurk samples are not themselves generalizable to all possible experiments.

They are also not necessarily generalizable to alternative online convenience sample marketplaces. For example, Coppock and McClellan (2019) explore and offer justification for the use of Lucid, a source of research subjects used in this paper. They conclude that "subjects recruited from the Lucid platform constitute a sample that is suitable for evaluating many social scientific theories." Coppock and McClellan replicate a series of survey experiments, demonstrating that the treatment effects estimated on high-quality nationally representative samples can be recovered using the Lucid sample.

In addition to this empirical result, the paper offers a useful discussion of the scientific value of convenience samples for testing theories. Coppock and McClellan accurately observe that concerns about external validity are often poorly motivated. Whether experiments conducted on a convenience sample and a nationally representative sample produce identical treatment effects is immaterial; what matters is whether a given sample is theoretically relevant. They propose, for example, that a sample of French speakers would serve as an inappropriate sample to test a theory about the effect of reading an English-language newspaper. *Literacy* is self-evidently a crucial moderator of textual media effects.

However, we would qualify the claim that "Whether or not future experiments will also exhibit low treatment effect heterogeneity is, of course, only a matter of speculation." We argue that digital literacy is likely to be a key moderator of digital media treatment effects, that it varies widely in the current U.S. population (and that it does so even among the current population of internet users), and that this is a significant problem for opt-in crowdworker platforms like MTurk for which digital literacy affects selection into the sample.

[10]The sociological processes by which political methodology deems certain elements of research design generally acceptable are central to the practice of political science but have largely escaped systematic study within the discipline. For example, it is possible to think of MTurk as a "trading zone" facilitating exchanges at the intersection of multiple fields and intellectual approaches (Collins, Evans, and Gorman 2007).

The trend of political life increasingly taking place online is not, we argue, "only a matter of speculation." Media technology has radically increased media heterogeneity, a development that shows no signs of abating. This is perhaps a problem for the ecological validity of a given media survey experiment, as a greater variety of stimuli are thus necessary to ensure a representative sample of even a specific type of media (Hovland, Lumsdaine, and Sheffield 1949). More immediately relevant is the continuing existence of the "second digital divide" in computer skills identified by Hargittai 20 years ago. In the absence of an unprecedented increase in the level of digital literacy among those at the bottom of the distribution, we have strong reason to expect that future experiments will exhibit greater effect heterogeneity along this dimension. But only, of course, if those experiments are conducted using samples with sufficient variation in the moderator.

8 Conclusion

As awareness increases of the potential for endogenous selection of digital literacy in commonly used samples, scholars will need to grapple with its substantive importance as a moderator in literatures not directly related to the study of online misinformation. Below we discuss three additional areas for consideration of effect heterogeneity by digital literacy.

Knowledge While political knowledge has long been established as a strong predictor of issue-position stability and ideological constraint, Kleinberg and Lau (2019) find that this relationship is now *reversed* for people belonging to the "internet generation." People for whom searching for information online is second nature do not need as much political information stored in their biological memories because they are able to access that information in their digital memories. These findings are consistent with psychological research that conceptualizes people's use of the internet as a kind of external or "transactive" memory (Sparrow, Liu, and Wegner 2011). Accordingly, people are better at remembering *where* to find information than the information itself if they expect to be able to find it later — even if this makes them think that they are more knowledgeable than they actually are (Fisher, Goddu, and Keil 2015).

Future research on the determinants and consequences of political knowledge will need to explicitly take into account variation in individuals' ability to seek out and process online information. Not only might the effect of knowledge vary across levels of digital literacy, but knowledge itself might function differently for people aware that answers will always be at their fingertips. Smith, Clifford, and Jerit (2020) suggest that "Scholars who are interested in measuring political knowledge should take efforts to minimize and diagnose search behavior," but this assumes a static, crystallized view of knowledge that may no longer hold for people who effectively externalize their store of factual information.

Nudges and defaults Behavioral scientists have demonstrated that classes of relatively low-cost interventions can have outside effects on behavior. These "nudges" have been promoted as a way to guide human decisions in a prosocial direction while preserving liberty of choice, as in the case of default options (Thaler and Sunstein 2009). Scholarship on the topic has focused on domains with a clear public welfare dimension, such as education and public health. But the insights of behavioral economics can be fruitfully applied to internet media, as contemporary accounts overlook the hidden obstacles and defaults that structure people's behavior. To take a simple example, modern web browsers come pre-loaded with bookmarks for large news and entertainment sites. Many people still use portals for email and other services which link to headlines, weather and other information. Sometimes, such sites automatically load on startup or with a new tab. It is not hard to customize one's settings, but the perceived cost of doing so may be too high for people lacking in digital literacy, with observable implications for political news diets (e.g., Guess 2019).

The effects of online choice architecture are not always innocuous, as more recent writing on "sludge" (Thaler 2018) and "dark patterns" (Mathur et al. 2019) has documented. As worries about online privacy and surveillance become more acute, the ability to protect oneself from insidious efforts to influence consumer and political choices will depend on what Hargittai and Micheli (2019) call "awareness of what is possible" — a dimension of internet skills comprising knowledge about default settings that can be changed. This arguably also covers defensive measures that can be taken against malicious or unscrupulous actors, such as installing blockers for illicit tracking. At a more basic level, people lacking in digital literacy may be unable to defend themselves against clickbait, misleading advertisements, and spam (perhaps political in nature; Perlstein 2012).

Microtargeting and campaign persuasion The possibility that digital literacy captures heretofore unobserved heterogeneity in people's responses to online "sludge" raises more general questions about the effects of modern campaign techniques. Evidence already suggests that online advertisements generate substantial heterogeneity; Kaptein and Eckles (2012) find that heterogeneity in social influence strategies is large relative to the average effect, for example. Moreover, since ads follow the same rules of "engagement" as other types of online content, campaign targeting of favored demographic and political subgroups is subject to feedback loops in which users most likely to respond to appeals will become even more likely to be targeted in the future via platforms' ad optimization algorithms (e.g., Eckles, Gordon, and Johnson 2018). Given generally low clickthrough rates, the kinds of people who are most likely to engage with commercial or political appeals on social media has important implications for democratic participation, representation, and campaign strategy. Future research should explore whether digital literacy explains variation in this important but understudied online political behavior.

References

Abbate, J. 2000. *Inventing the Internet.* Inside technology M I T PRESS (MA).
URL: *https://books.google.ch/books?id=E2BdY6WQo4AC*

Arceneaux, Kevin, and Martin Johnson. 2013. *Changing minds or changing channels?: Partisan news in an age of choice.* University of Chicago Press.

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.

Bawden, David. 2008. "Origins and Concepts of Digital Literacy." In *Digital Literacies: Concepts, Policies and Practices*, ed. Colin Lankshear, and Michele Knobel. pp. 17–32.

Bennett, W Lance, and Shanto Iyengar. 2008. "A new era of minimal effects? The changing foundations of political communication." *Journal of communication* 58 (4): 707–731.

Berinsky, Adam J, Gregory A Huber, and Gabriel S Lenz. 2012. "Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk." *Political analysis* 20 (3): 351–368.

Bond, Robert M, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler. 2012. "A 61-million-person experiment in social influence and political mobilization." *Nature* 489 (7415): 295.

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.

Burke, P.C.H.F.E.C.P., A. Briggs, P. Burke, and U.P. Burke. 2005. *A Social History of the Media: From Gutenberg to the Internet.* Wiley.
URL: *https://books.google.ch/books?id=OoxsKnfVzLwC*

Chandler, Jesse, Pam Mueller, and Gabriele Paolacci. 2014. "Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers." *Behavior research methods* 46 (1): 112–130.

Clifford, Scott, and Jennifer Jerit. 2016. "Cheating on political knowledge questions in online surveys: An assessment of the problem and solutions." *Public Opinion Quarterly* 80 (4): 858–887.

Collins, Harry, Robert Evans, and Mike Gorman. 2007. "Trading zones and interactional expertise." *Studies in History and Philosophy of Science Part A* 38 (4): 657–666.

Converse, Phillip. 1964. "The Nature of Belief Systems in Mass Publics. In Ideology and Discontent, ed. David Apter. New York: Free Press.".

Coppock, Alexander. 2019. "Generalizing from survey experiments conducted on Mechanical Turk: A replication approach." *Political Science Research and Methods* 7 (3): 613–628.

Coppock, Alexander, and Oliver A McClellan. 2019. "Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents." *Research & Politics* 6 (1): 2053168018822174.

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.

Dunaway, Johanna, Kathleen Searles, Mingxiao Sui, and Newly Paul. 2018. "News attention in a mobile era." *Journal of Computer-Mediated Communication* 23 (2): 107–124.

Eckles, Dean, Brett R. Gordon, and Garrett A. Johnson. 2018. "Field studies of psychologically targeted ads face threats to internal validity." *Proceedings of the National Academy of Sciences* 115 (23): E5254–E5255.
URL: *https://www.pnas.org/content/115/23/E5254*

Eshet, Yoram. 2004. "Digital literacy: A conceptual framework for survival skills in the digital era." *Journal of educational multimedia and hypermedia* 13 (1): 93–106.

Fisher, Matthew, Mariel K Goddu, and Frank C Keil. 2015. "Searching for explanations: How the Internet inflates estimates of internal knowledge." *Journal of experimental psychology: General* 144 (3): 674.

Flanagin, Andrew J, and Miriam J Metzger. 2007. "The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information." *New Media & Society* 9 (2): 319–342.

Frederick, Shane. 2005. "Cognitive reflection and decision making." *Journal of Economic Perspectives* 19 (4): 25–42.

Gilster, Paul. 1997. *Digital Literacy*. Wiley Computer Pub. New York.

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. "Fake news on Twitter during the 2016 US presidential election." *Science* 363 (6425): 374–378.

Guess, Andrew, Jonathan Nagler, and Joshua Tucker. 2019. "Less than you think: Prevalence and predictors of fake news dissemination on Facebook." *Science Advances* 5 (1): eaau4586.

Guess, Andrew M. 2015. "Measure for Measure: An Experimental Test of Online Political Media Exposure." *Political Analysis* 23 (1): 59–75.

Guess, Andrew M. 2019. "(Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets." Working paper available at `https://www.dropbox.com/s/4qbn29l0pkpe2oa/Guess_OMD.pdf?dl=0`.

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

Hargittai, Eszter. 2005. "Survey measures of web-oriented digital literacy." *Social science computer review* 23 (3): 371–379.

Hargittai, Eszter. 2009. "An update on survey measures of web-oriented digital literacy." *Social science computer review* 27 (1): 130–137.

Hargittai, Eszter, and Aaron Shaw. 2020. "Comparing Internet Experiences and Prosociality in Amazon Mechanical Turk and Population-Based Survey Samples." *Socius* 6: 2378023119889834.

Hargittai, Eszter, and Marina Micheli. 2019. "Internet skills and why they matter." *Society and the Internet: How Networks of Information and Communication are Changing Our Lives* p. 109.

Hovland, Carl I, Arthur A Lumsdaine, and Fred D Sheffield. 1949. "Experiments on mass communication.(Studies in social psychology in World War II), Vol. 3.".

Kahne, Joseph, Nam-Jin Lee, and Jessica Timpany Feezell. 2012. "Digital media literacy education and online civic and political participation." *International Journal of Communication* 6: 24.

Kaptein, Maurits, and Dean Eckles. 2012. "Heterogeneity in the Effects of Online Persuasion." *Journal of Interactive Marketing* 26 (3): 176 – 188.
URL: *http://www.sciencedirect.com/science/article/pii/S1094996812000035*

Kleinberg, Mona S, and Richard R Lau. 2019. "The Importance of Political Knowledge for Effective Citizenship: Differences Between the Broadcast and Internet Generations." *Public Opinion Quarterly* 83 (2): 338–362.

Knox, Dean, Will Lowe, and Jonathan Mummolo. Forthcoming. "The Bias Is Built In: How Administrative Records Mask Racially Biased Policing." *American Political Science Review* .

Koltay, Tibor. 2011. "The media and the literacies: Media literacy, information literacy, digital literacy." *Media, Culture & Society* 33 (2): 211–221.

Luskin, Robert C. 1987. "Measuring political sophistication." *American journal of political science* pp. 856–899.

Luskin, Robert C. 1990. "Explaining political sophistication." *Political Behavior* 12 (4): 331–361.

Mathur, Arunesh, Gunes Acar, Michael J Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. "Dark patterns at scale: Findings from a crawl of 11K shopping websites." *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW): 1–32.

Metzger, Miriam J, Andrew J Flanagin, Alex Markov, Rebekah Grossman, and Monica Bulger. 2015. "Believing the unbelievable: understanding young people's information literacy beliefs and practices in the United States." *Journal of Children and Media* 9 (3): 325–348.

Munger, Kevin. 2019. "Temporal Validity in Online Social Science.".

Munger, Kevin, Andrew Guess, Jonathan Nagler, and Joshua A Tucker. 2018. "How Accurate Are Survey Responses on Social Media and Politics?" *Political Communication* .

Munger, Kevin, Mario Luca, Jonathan Nagler, and Joshua Tucker. 2020. "The (Null) Effects of Clickbait Headlines On Polarization, Trust, and Learning." *Public Opinion Quarterly* Forthcoming.

Mutz, Diana C. 2015. *In-your-face politics: The consequences of uncivil media*. Princeton University Press.

Osmundsen, Mathias, Alexander Bor, Peter Bjerregaard Vahlstrup, Anja Bechmann, and Michael Bang Petersen. 2020. "Partisan polarization is the primary psychological motivation behind "fake news" sharing on Twitter.".

Pearl, Judea. 2009. *Causality*. Cambridge university press.

Pennycook, Gordon, and David G Rand. 2019. "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning." *Cognition* 188: 39–50.

Pennycook, Gordon, James Allan Cheyne, Paul Seli, Derek J Koehler, and Jonathan A Fugelsang. 2012. "Analytic cognitive style predicts religious and paranormal belief." *Cognition* 123 (3): 335–346.

Pennycook, Gordon, Robert M Ross, Derek J Koehler, and Jonathan A Fugelsang. 2016. "Atheists and agnostics are more reflective than religious believers: Four empirical studies and a meta-analysis." *PloS one* 11 (4).

Perlstein, Rick. 2012. "The Long Con: Mail-order conservatism." *The Baffler* (21).
URL: *https://thebaffler.com/salvos/the-long-con*

Prior, Markus. 2007. *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.

Redmiles, Elissa M, Neha Chachra, and Brian Waismeyer. 2018. Examining the Demand for Spam: Who Clicks? In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM p. 212.

Robinson, L., Rima Kupryte, Peter Burnett, and David Bawden. 2000. "Libraries and the Internet: overview of a multinational training course." *Program* 34 (2): 187–194.

Sharkey, Jennifer. 2013. "Establishing Twenty-First-Century Information Fluency." *Reference & User Services Quarterly* 53: 33–39.

Smith, Brianna, Scott Clifford, and Jennifer Jerit. 2020. "TRENDS: How Internet Search Undermines the Validity of Political Knowledge Measures." *Political Research Quarterly* 73 (1): 141–155.
URL: *https://doi.org/10.1177/1065912919882101*

Sood, Gaurav, and Yphtach Lelkes. 2018. "Don't Expose Yourself: Discretionary Exposure to Political Information." In *Oxford Research Encyclopedia of Politics*.

Sparrow, Betsy, Jenny Liu, and Daniel M. Wegner. 2011. "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips." *Science* 333 (6043): 776–778.
URL: *https://science.sciencemag.org/content/333/6043/776*

Stagnaro, Michael N, Gordon Pennycook, and David G Rand. 2018. "Performance on the Cognitive Reflection Test is stable across time." *Judgment and Decision Making* 13 (3): 260–267.

Sundar, S Shyam, and Sampada S Marathe. 2010. "Personalization versus customization: The importance of agency, privacy, and power usage." *Human Communication Research* 36 (3): 298–322.

Thaler, Richard H. 2018. "Nudge, not sludge." *Science* .

Thaler, Richard H, and Cass R Sunstein. 2009. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Vraga, Emily K, and Melissa Tully. 2015. "Media literacy messages and hostile media perceptions: Processing of nonpartisan versus partisan political information." *Mass Communication and Society* 18 (4): 422–448.

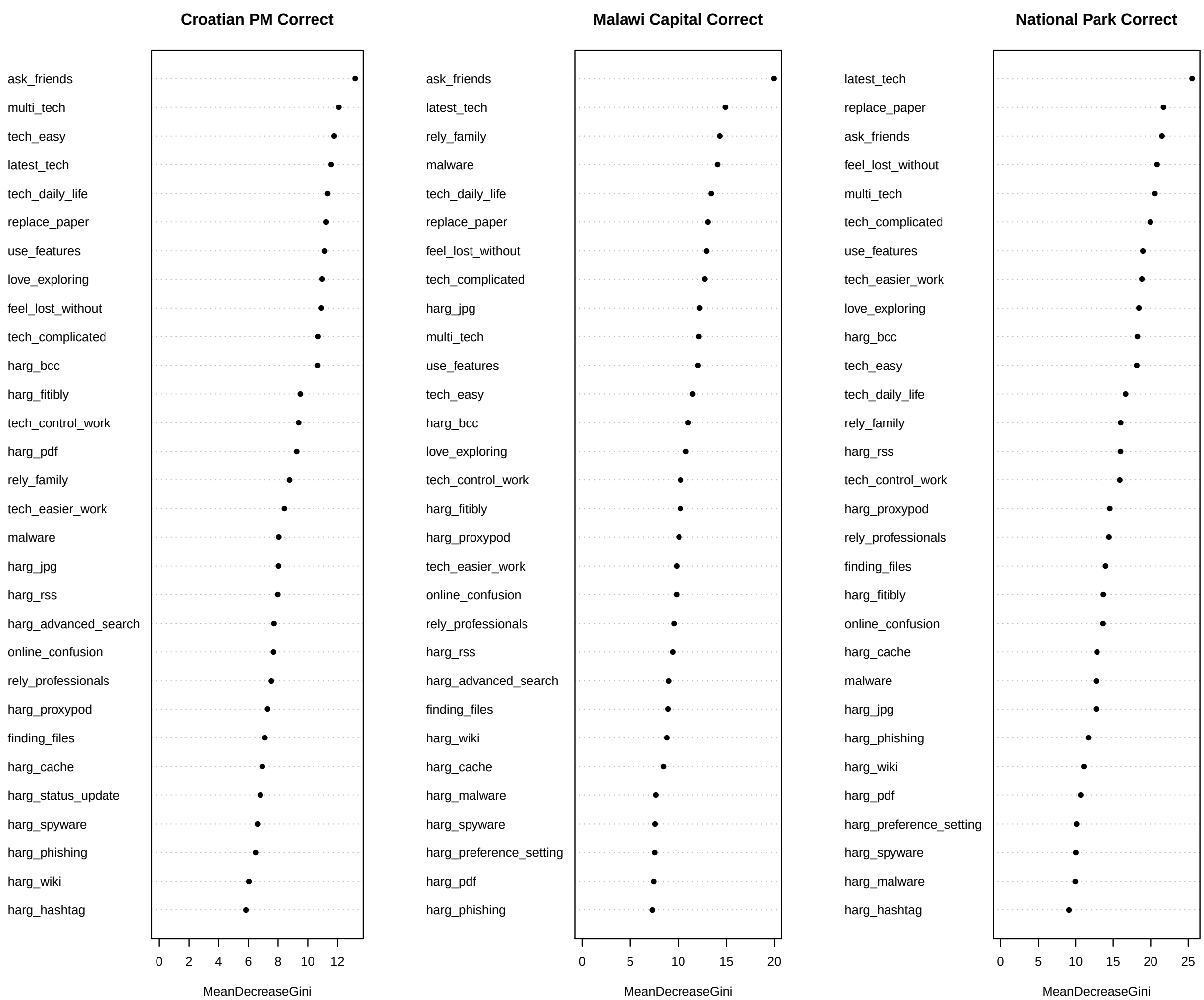
Vraga, Emily K, Melissa Tully, John E Kotcher, Anne-Bennett Smithson, and Melissa Broeckelman-Post. 2015. "A multi-dimensional approach to measuring news media literacy." *Journal of Media Literacy Education* 7 (3): 41–53.

Wineburg, Sam Author, and Sarah McGrew. 2018. "Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information." *sociology* 2015: 12–15.

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.

Appendix

Figure 8: Increased Classification Accuracy of Information Retrieval, Aggregated Measures



Survey Questions

Bolded items are selected for inclusion in our validation procedure.

"Digital Literacy" Scale (adapted from Hargittai 2009)

How familiar are you with the following computer and Internet-related items? Please choose a number between 1 and 5 where 1 represents "no understanding" and 5 represents "full understanding" of the item.

- **Phishing**
- Preference Setting
- App
- **Hashtag**
- Social Media
- Status Update
- Spyware
- Selfie
- Wiki
- Advanced Search
- PDF

Next screen:

- Tagging
- Tablet
- Smartphone
- **JPG**
- **Malware**
- **Cache**
- BCC (on email)
- **RSS**
- Proxypod
- Fitibly

(last two items are intentionally made up)

Power User Scale (adapted from Sundar and Marathe 2010)

Please indicate your agreement with the following statements on a scale of -4 = Strongly Disagree to 4 = Strongly Agree.

- I think most technological gadgets are complicated to use.
- I make good use of most of the features available in any technological device.
- I have to have the latest available upgrades of the technological devices that I use.
- Use of information technology has almost replaced my use of paper.
- I love exploring all the features that any technological gadget has to offer.
- I often find myself using many technological devices simultaneously.

Next screen:

- **I prefer to ask friends how to use any new technological gadget instead of trying to figure it out myself.**
- Using any technological device comes easy to me.
- **I feel like information technology is a part of my daily life.**
- Using information technology gives me greater control over my work environment.
- **Using information technology makes it easier to do my work.**
- I would feel lost without information technology.

Low End Scale

Please indicate how often these statements apply to you. [Never / Almost never / Occasionally / Somewhat often / Very often]

- I rely on family members to introduce me to new technology.
- I have professionals (such as the Geek Squad) or family members take a look at my
- computer when something isn't working.
- A lot of the things I see online confuse me.
- I have problems with viruses and malware on my computer.
- **I have trouble finding things that I've saved on my computer.**