Exposure to campaign claims in mobile browsing and survey data

Susan Banducci, Laszlo Horvath, Dan Stevens
University of Exeter

Working paper, presented at the 116th American Political Science Association's Annual Meeting, 10 September 2020

1st September 2020

I BACKGROUND

It has been established again and again since Converse (1964) that levels of knowledge about politics and policy in democracies are low (e.g. Carpini and Keeter, 1996). What has changed is whether such levels of ignorance warrant the inference that individuals are therefore unable to hold incumbent governments to account. Views of what members of the public need to know have shifted (e.g. Hochschild and Einstein, 2015), along with evidence that individuals need only draw on limited amounts of cognitive and affective information to reach accurate conclusions (Sniderman et al., 1991). In addition, individuals appear able to update their beliefs with new information (Luskin et al., 2002), while collective public opinion also responds rationally to change (Page and Shapiro, 2010).

There are a number of concerns to highlight, however. First, for even if individuals can make accurate inferences from limited information, this still relies upon the information not being systematically skewed (e.g. Jerit and Barabas, 2012). Second, individuals may simply see too much misinformation and disinformation to be able to make accurate inferences (Jerit and Zhao, 2020). Recently, research has shown the prevalence of misinformation originating from mainstream political actors (Swire et al., 2017) and spread during election campaigns.

We ask how citizens respond to repeated exposure to these messages from media sources. Drawing on a two-wave panel study capturing the full log of mobile user activity, we link a behavioural measure of information exposure to survey data on the believability of eight fact-checked claims that were spread during the 2017 UK General Election campaign. We find that the effect of media exposure on the over-time adjustment of beliefs varies depending on the claim. Exposure in three cases led to the adjustment of beliefs if it came from sources congruent with the broader political narrative around these claims, whether correct or incorrect. We find no evidence that balanced media exposure led to correcting beliefs in false statements overall.

II HYPOTHESES

Drawing on processing fluency theory and its applications to political communication research (Oppenheimer, 2008; Berinsky, 2017), we hypothesise that greater exposure to campaign claims increases the ease with which information is recalled and thus the likelihood that citizens believe them. In the news media, campaign claims are likely to be embedded in a broader narrative which may be consistent with the original claim, critical to the claim, or a mixture of both. We thus specify that source congruence is necessary for heuristic fluency to work in practice:

> **Source congruence hypothesis.** Exposure to claims from sources consistent with the claims' original narrative is more likely to strengthen beliefs than exposure from sources inconsistent with the narrative.

We then ask about the conditions necessary for citizens to *correct* pre-existing false beliefs. Exposure to an uncontested political narrative is unlikely to do so (e.g. British partisan press or American cable news), we thus hypothesise that exposure to a balance across multiple sources presenting a variety of narratives matters when it comes to correcting misperceptions. In addition, recent research underscores the importance of perceived source quality to counter misinformation (Pennycook and Rand, 2019) thus exposure to the claims from impartial sources (e.g. British broadcast news or fact checking information sites) should also lead to corrections.

> **Balance hypothesis.** Balanced exposure to claims from both congruent and incongruent sources, or from sources that are likely to present balanced or impartial coverage, is likely to correct misperceptions.

III SURVEY METHODS

Sample and attrition

The surveys were administered online by ICM unlimited. The first wave of surveys were completed by May 28th—and re-interviewed after the election from June 9 to June 13. Our final sample size is subject to two-fold attrition as well as missing data: of the 2,523 who took wave 1 of the survey, 1,841 were re-interviewed, 1,072 opted into the tracking component (see next sections), and finally 820 respondents answered our survey measures of interest. The demographic quotas of both surveys were managed by the data provider to mirror the UK's 2011 census in terms of age, gender, and current geographic location.

Survey measures

Our repeated measures dependent variable is respondents' belief in a set of eight campaign claims, each expressed on a five-point scale ranging from "strongly disbelieve" to "strongly believe." Table 1 on page 3 gives an overview of these claims, showing which political positions they are congruent with (leave versus remain referendum vote or pro versus anti-Conservative government), and whether factually correct. Figure 1 shows the dynamics of these beliefs across the two waves. In addition, we asked about familiarity with the claim in wave 1, whether respondents have heard them "never," "not very often," "quite often," or "very often." We use this as a predictor of baseline beliefs, as a self-reported measure of fluency. We then proceed to investigating what explains the over-time dynamics of beliefs, tracking mobile user activity.

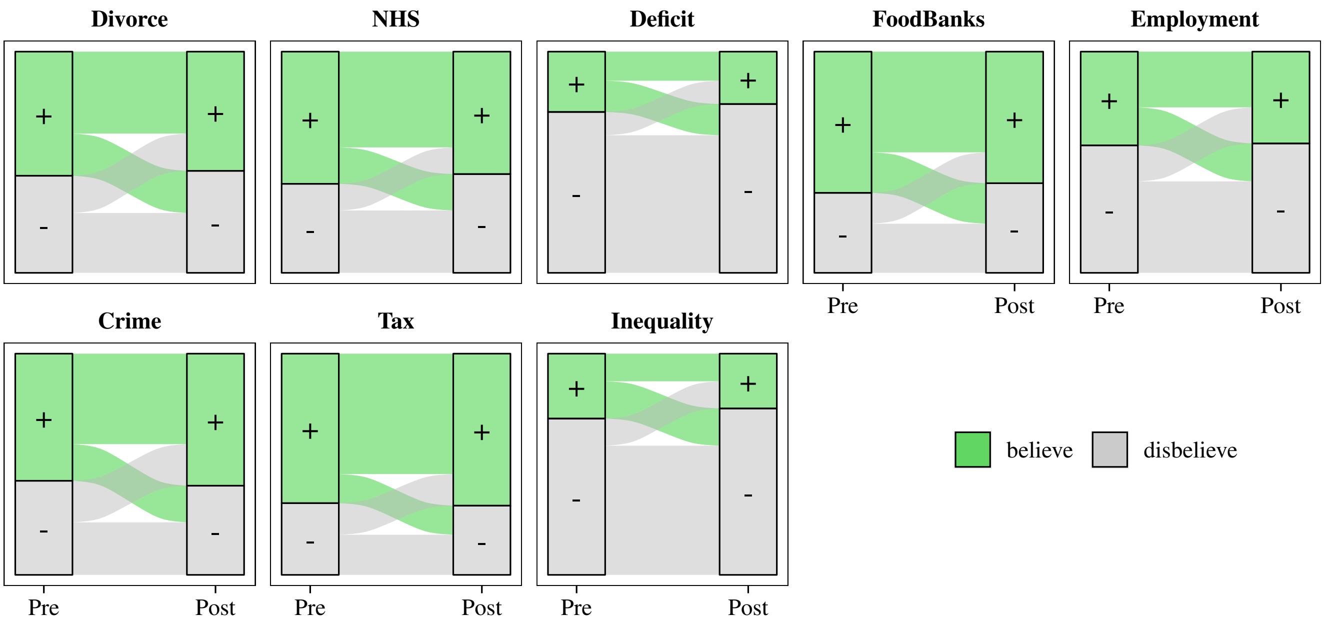

Figure 1: Over-time dynamics of beliefs in campaign claims

IV MOBILE USER TRACKING

Tracking data

The tracking data, provided to us by ICM, contains 14,876,933 URLs, which is the full log of client-server communication on 1,072 respondents' mobile devices, relating to activity on a large number of applications including mobile browsers, news reading apps, and social media apps. Being an opt-in component of the study, the time span of tracking itself varies across these respondents from 6.5 hours to fifteen days, as shown in Figure 2 for a sample of respondents. We address this issue standardising news exposure over the duration of tracking for each respondent in our models. Our challenge is to identify the small subset of requests that point to news articles about the 2017 General Election, to extract the full text of these articles, and to identify which ones featured either of the eight campaign claims above. We approach this problem using a number of methods.

Inductive approach

We first identified the URLs that point to a *domain* that is classified as a "news domain" by Amazon's Alexa.com database (we extracted UK local, national, and global news domains). We found 37,652 such records. The vast majority of these do not point to news articles, however, but to segments of these websites including images, videos, tracking cookies, and other applications. We thus requested further information about these URLS using Diffbot's News API that returned, if linked to a valid news story, full text and other contextual information in English. With this method, we obtained 2,642 stories, accessed by 434 users.

Merging with online news archives

To complement this, we then referred to external online news databases that archive URL data from news and information websites. We requested archival data from Webhose API news,

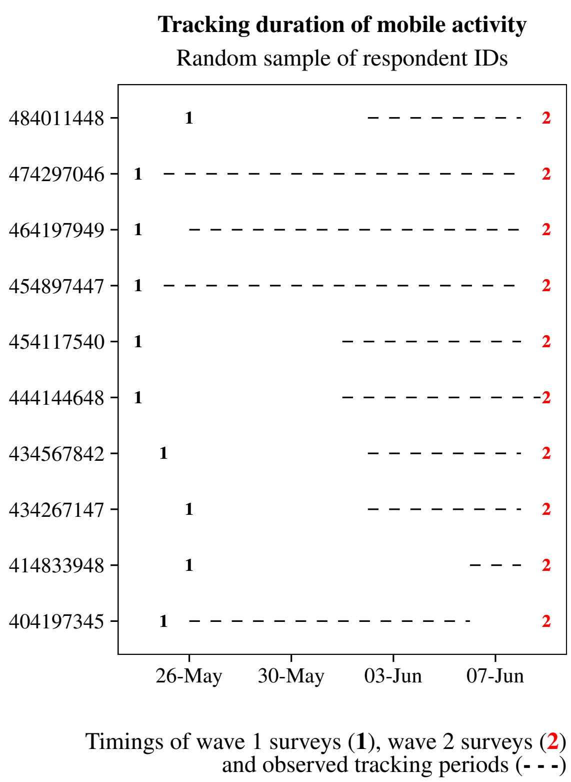

Figure 2: Tracking duration of mobile activity

Table 1: Overview of campaign claims

Claim ID	Claim	Congruence	Correct
Divorce	The EU wants the UK to pay £50-60 billion before they negotiate a post-Brexit trade deal.	Leave +	No*
NHS	The NHS is under unprecedented pressure due to an influx of EU migrants that has forced doctors to take on 1.5 million extra patients in three years.	Leave +	No
Deficit	The deficit has fallen by two-thirds since 2010.	Conservative +	Yes
FoodBanks	The number of people using food banks since 2010 has gone from the tens of thousands to the millions.	Conservative -	No
Employment	The number of people in employment in the UK is historically high, as is the proportion of people in work.	Conservative +	Yes
Crime	Almost every police force in the country recorded an increase in crime over the last year.	Conservative -	Yes
Tax	The Treasury loses £40 billion each year due to tax evasion and avoidance by the super-rich and corporate elites.	Conservative -	No
Inequality	Income inequality is narrowing in the UK.	Conservative +	No

*Full Fact were unable to verify this claim because of lack of information, rather than factual incorrectness."

which crawled 40,478 news stories relating to the 2017 General Election, published during the tracking period. This matched an additional 874 records, thus the final text corpus is 3,516, accessed by 444 respondents.

Identify claims in corpus

We treated claim detection as an NLP semantic similarity problem. The task was to obtain a metric that captures, on the one hand, the meaning of the 87,910 sentences that occur in the text corpus above—the *target sentences*, and on the other, the meaning of our set of eight campaign claims—the *reference sentences*; the aim being to calculate the distance (similarity) between these two. In practice, rather than using the one sentence claim as asked in the surveys as the reference sentence, we seeded more reference sentences for each claim using simple a keyword search in the news corpus, to better capture the vocabulary surrounding topic (for example, we also seeded sentences using the phrase "divorce bill" instead of "paying £50-60 billion before they negotiate post-Brexit trade deal," as worded in the survey). To calculate similarity, We used GloVe (Pennington et al., 2014) word embeddings implemented in spaCy. GloVe represents each word as a large multi-dimensional vector in the context of the sentence it occurs, which is used as a general-purpose solution for semantic similarity.

We used a similarity score in the top 99.9th percentile to conclude if a target sentence comes close enough to the reference sentences. We observe that rather than capturing the claims literally, we get sentences that relate to the topic to the claim but may present it in a differently in a variety of context. For example, in relation to the first campaign claim about the £50-60bn post-Brexit trade deal, we obtain matching sentences that mention "the astronomical size of a proposed divorce bill from Brussels which would reach £85 billion"—a very close match, but also "Brussels forcing Britain to pay into their system is 'wholly without merit in law'" which doesn't mention a Brexit bill sum but does relate to the broader topic of the claim. An example of a false positive is a claim around the financial costs of the US-Mexico wall, an issue mentioned in our corpus. We finally aggregate these claims on the whole document level which is coded 1 if featuring any number of target sentences similar to the reference claim, and 0 if none. We found a total of 528 articles that feature either of the eight reference claims, 15% of all records.

Adjusting exposure to source congruence

Before collapsing the browsing data on the individual level, we examined the 52 sources where the claims were featured for congruence with the original narrative around the claim. We show these in Table 1's column "Congruence". To determine if a source endorsed either side of the Leave campaign in 2016, which is relevant for the first two claims, we refer to the University of Oxford's Reuters Institute for the Study of Journalism (2016) report, and coded -1 if source backed Remain, +1 if source backed Leave, and 0 for neutral or balanced sources, such as all broadcast news websites. On the individual-level, then, a respondent's "source congruence" is the average of her viewed sources' Leave endorsement. Similarly, we examined sources for congruence with endorsing the Conservative campaign in 2017, which is available from the major sources' content analysed in the University of Loughborough's Centre for Research in Communication and Culture (2017) report or, if not available, we checked their editorial statement during the 2017 campaign. When adjusting exposure to congruence, we multiply the number of times news articles were viewed with the congruence score, with greater values indicating exposure to the claim along with its original narrative on a congruent source, while negative values indicate that the claim was accessed on incongruent sources, thus likely contested.

Adjusting exposure to balance

In the scale established above, values near zero indicate either a balance of exposure to both incongruent and congruent messages, or that exposure to the claim was directly from sources that were coded "balanced," for example the impartial broadcast news coverage or newspapers that did not endorse either parties in 2017 or took no position during the Brexit referendum in 2016. To be able to test the hypothesis that balanced coverage may lead to learning and thus correcting beliefs, we invert the congruence scale where higher values indicate exposure to a balance of messages or to balanced sources, and negative values indicate exposure on either congruent or incongruent sources.

V RESULTS

Congruent exposure effects

We first fit eight separate linear models for each claim, regressing wave 2 belief on predisposition (wave 1 belief), familiarity in terms of how often respondents' recall hearing about the claim, exposure in tracking data, adjusted to tracking duration (number of stories featuring claims viewed divided by duration

Table 2: Unpooled models

	Divorce	NHS	Deficit	Food	Employment	Crime	Tax	Inequality
	Dependent variable: Believe claim (1–5)							
Intercept	2.50^{***}	1.55^{***}	1.28^{***}	1.68^{***}	1.90^{***}	2.70^{***}	2.32^{***}	1.21^{***}
	(0.14)	(0.14)	(0.10)	(0.14)	(0.13)	(0.14)	(0.14)	(0.10)
Familiarity with claim	0.08^{**}	0.00	0.12^{***}	0.16^{***}	0.08^{**}	0.04	0.05	0.17^{***}
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Wave 1 belief in claim	0.23^{***}	0.53^{***}	0.44^{***}	0.40^{***}	0.34^{***}	0.25^{***}	0.36^{***}	0.35^{***}
	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
Exposure adjusted to congruence	0.00	0.11^{***}	0.59	0.02	0.10^{*}	0.02	-0.02	0.09^{**}
	(0.04)	(0.04)	(0.37)	(0.03)	(0.06)	(0.05)	(0.03)	(0.04)
R^2	0.07	0.27	0.27	0.21	0.14	0.07	0.15	0.21
Adj. R^2	0.06	0.27	0.26	0.20	0.13	0.06	0.14	0.21
Num. obs.	820	820	820	820	820	820	820	820
RMSE	0.97	1.08	0.87	1.01	0.98	0.96	0.89	1.04

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

in seconds) and to source congruence thus greater values indicate evidence of exposure to the original claim from sources consistent with the original narrative. These models are presented in Table 2. We find differences across these claims in terms of how much of the over-time believability is explained by exposure. For five of our eight claims, we are unable to discern an effect of news exposure. For the other three claims, we find exposure effects on the (incorrect) NHS statement, the (incorrect) narrowing inequality statements and the (correct) employment statement. In these cases, respondents were more likely to adjust their belief if reading about these claims from sources that were likely to provide a narrative consistent with the claim, rather than challenging it. We did not hypothesise a mechanism that would explain these differences across the claims. We do note, however, that the initial aggregate levels of belief were low for these three claims and thus most open for adjustment, potentially via media effects but also other mechanisms. Pooling all claims together in a mixed effects model, as shown in the left column of Table 3, we find a small overall effect of exposure, accounting for individual-level, between-claims, and over-time variances. We also note that, of these second-level variance components, variation across claims is the largest, suggesting future work to conceptualise differences about claim.

Table 3: Pooled models

	Believe claim	Correct about claim
Intercept	3.39^{***}	2.80^{***}
	(0.14)	(0.20)
Familiarity with claim	0.35^{***}	-0.11^{***}
	(0.01)	(0.01)
Exposure adjusted to congruence	0.05^{***}	
	(0.01)	
Exposure adjusted to balance		0.00
		(0.01)
AIC	37210.73	38952.38
BIC	37263.10	39004.75
Log Likelihood	-18598.36	-19469.19
N	13120	13120
N respondents	820	820
N claims	8	8
N waves	2	2
Variance comp. respondents	0.11	0.01
Variance comp. claims	0.14	0.30
Variance comp. waves	0.00	0.00
Variance comp. residual	0.93	1.12

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

Balanced exposure effects

The second column of Table 3 shows our model exploring the effect of balanced exposure. We first recoded the dependent variable so that higher values indicate holding correct beliefs, regardless of congruence. We also inverted the exposure scale so that higher values indicate either a balance between congruent and incongruent messages, or that exposure is from sources that provide balanced coverage of issues (e.g. British broadcast news). We fail to detect an effect of exposure in this sense, however. Interestingly, the self-reported exposure variable taken at Wave 1, "familiarity" shows negative association thus respondents who heard about the claim more often and thus more fluent were also likely to hold incorrect beliefs.

VI DISCUSSION

Looking at the over-time dynamics of beliefs in eight campaign claims, we found some evidence of the heuristic fluency process in that repeated exposure led to believing the claims. For five claims, self-reported previous exposure, "familiarity" already meant believing them in wave 1; and for a subset of three claims, we found additional effects over-time and within-subjects, using a behavioural measure of news exposure, provided that the claims were featured on a source congruent with the political narrative the original claim. We found no evidence that exposure to a balance of sources, or sources providing impartial coverage, led to correcting misperceptions.

REFERENCES

Berinsky, A. J. (2017). Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*, 47(2):241–262.

Carpini, M. X. D. and Keeter, S. (1996). *What Americans know about politics and why it matters*. Yale University Press.

Centre for Research in Communication and Culture (2017). Newspapers remain hostile to Labour in their election coverage. https://www.lboro.ac.uk/media-centre/press-releases/2017/may/newspapers-remain-hostile-to-labour-in-coverage.

Converse, P. E. (1964). The nature of belief systems in mass publics. *Critical Review*, 18(1-3):1–74.

Hochschild, J. L. and Einstein, K. L. (2015). *Do facts mat-*

ter?: Information and misinformation in American politics, volume 13. University of Oklahoma Press.

Jerit, J. and Barabas, J. (2012). Partisan perceptual bias and the information environment. *The Journal of Politics*, 74(3):672–684.

Jerit, J. and Zhao, Y. (2020). Political misinformation. *Annual Review of Political Science*, 23:77–94.

Luskin, R. C., Fishkin, J. S., and Jowell, R. (2002). Considered opinions: Deliberative polling in britain. *British Journal of Political Science*, pages 455–487.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in cognitive sciences*, 12(6):237–241.

Page, B. I. and Shapiro, R. Y. (2010). *The Rational Public: Fifty years of trends in Americans' policy preferences*. University of Chicago Press.

Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Pennycook, G. and Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, 116(7):2521–2526.

Reuters Institute for the Study of Journalism (2016). UK newspapers' positions on Brexit. https://www.ox.ac.uk/news/2016-05-23-uk-newspapers-positions-brexit.

Sniderman, P. M., Brody, R. A., and Tetlock, P. E. (1991). The role of heuristics in political reasoning: A theory sketch. In *Reasoning and Choice: Explorations in Political Psychology*, pages 14–30. Cambridge University Press.

Swire, B., Berinsky, A. J., Lewandowsky, S., and Ecker, U. K. (2017). Processing political misinformation: comprehending the trump phenomenon. *Royal Society Open Science*, 4(3):160802.