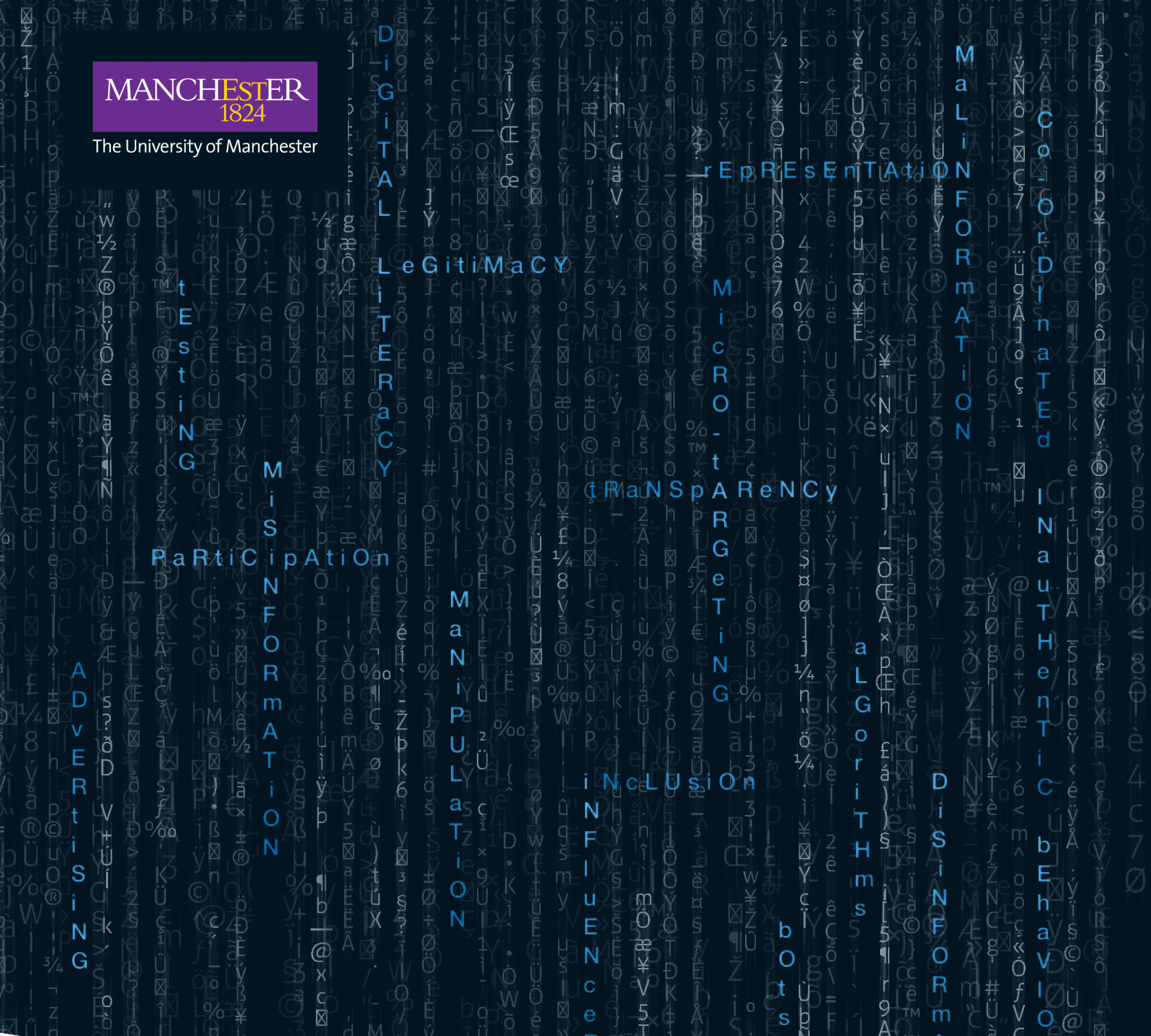


DEMOCRACY@RISK

Democracy in the era of digital micro-targeting and online misinformation

By ARIADNA TSENINA, RACHEL GIBSON, EMMA BARRETT

October 2021

About

Democracy@Risk is a research project led by Professor Emma Barrett and Professor Rachel Gibson at the University of Manchester. The project is focused on identifying, measuring and responding to the new threats (and opportunities) that the deployment of digital technologies and AI pose to the functioning of democratic institutions and the deeper cultural norms of trust that underpin them. Democracy@Risk reviews the existing knowledge base from an interdisciplinary perspective, working with policy-makers and researchers across the UK and internationally to establish priority areas for new research, evidence and policy.

Acknowledgements

This research was made possible thanks to the generous funding from the University of Manchester Research Institute Pump Priming Programme.

Whilst all errors are the authors' own, this report has benefited greatly from the time and expertise contributed generously by a large number of scholars and policy-makers committed to protecting democracy in the face of new challenges posed by digital communication and technology.

We are especially grateful to Dr Kate Dommett, Professor Martin Innes and Dr Kari Kivinen for their insightful, constructive and detailed comments on this work, which have been particularly instrumental in the formulation of the report's final recommendations on the next steps for policy-oriented research in this field.

Our work owes much to the scholars who have attended our three academic workshops in May 2020 - Dr Julio Amador Diaz Lopez, Dr Nick Anstead, Oliver Beatson, Professor Colin Bennett, Dr Esmeralda Bon, Professor Pete Burnap, Dr Kate Dommett, Professor Karen Douglas, Dr Arne Hintz, Professor Martin Innes, Professor Adam Joinson, Dr Rachel Hoffman, Professor Stephen Hutchings, Dr Vitaly Kazakov, Dr Catriona Kennedy, Dr Sanne Kruikemeier, Dr Simon Kruschinski, Professor Stephan Lewandowsky, Dr Cerwyn Moore, Professor Victoria Nash, Professor Sarah Oates, Professor Andrea Römmele, Dr David Schoch, Professor Vera Tolz, Dr Rebekah Tromble, Professor Cristian Vaccari, Dr Geoff Walton, and Professor Dominic Wring. Our team learned a lot from these discussions and we are grateful to these researchers for taking the time out of their busy schedules to share expertise on the core themes of the report, provide us with feedback and recommend additional literature for the review.

We would like to express our sincere gratitude to the scholars and policy-makers who attended the Policy Forum we held in partnership with the Alan Turing Institute in February 2020. These discussions played a fundamental role in establishing the initial directions for our project.

Finally, we would also like to thank the Digital Futures team, led by Sarah Barton, at the University of Manchester for their hard work and continued support for our project.

Contents

Executive Summary 1

Introduction 3

Digital Political Micro-Targeting 4

How does it work? 4

How new is it? 5

Can digital micro-targeting 'steal' elections through subliminal persuasion? 7

Does digital political micro-targeting receive adequate oversight? 10

Does digital micro-targeting entrench problematic patterns in political behaviour? 11

Recommendations 13

Online Misinformation 14

Who creates online misinformation? 15

How does misinformation grow? 16

Online misinformation as an eco-system 17

Why is misinformation harmful for democracy? 17

Can misinformation be corrected? 19

Can we 'inoculate' against misinformation? 21

Recommendations 22

Digital Information Literacy 24

What is Digital Information Literacy? 24

Characteristics of DIL as a solution 25

DIL in education 25

DIL education in the UK 26

Recommendations 29

Conclusion 30

Next Steps: A Note for Researchers 31

Appendix 32

References 33

Executive Summary

This report summarises the key findings of a multidisciplinary literature review conducted by the Democracy@Risk Project at the University of Manchester in 2020. Drawing on existing research and scholarship, it sheds further light on *digital political micro-targeting* and *online misinformation* as two potential sources of harm for democracy in the digital age, and assesses the challenges relating to *digital information literacy* as one of the most frequently-proposed solutions to the problems generated by these phenomena.

Digital Political Micro-Targeting

- Digital political micro-targeting suffers from **severe gaps in oversight** as a result of outdated electoral legislation and ambiguities in data protection regulation, lack of industry standards for record-keeping and reporting, use of opaque machine-learning technologies, and poor controls available to individual users.

- This practice also **incentivises and facilitates problematic political behaviour**, including political redlining, voter suppression, fragmentation of policy promises and emphasis of wedge issues.

- The above problems generate **considerable risks for democracy**, which include the reduction of accountability, lower levels of participation, distorted representation, and erosion of public trust in democratic institutions and political actors.

Online Misinformation

- Online misinformation can be understood as a **complex eco-system** comprising a wide range of actors, motivations, outlets for publication, and pathways for spreading problematic content

- Misinformation **affects political behaviour** by reinforcing political beliefs, eroding trust, encouraging political apathy, and increasing polarisation.

- These effects carry **negative implications for democracy**, threatening to entrench harmful political attitudes, erode faith in political actors and institutions, weaken the impact of government guidance, reduce political participation, inhibit the formation of collective will and undermine the legitimacy of democratic decisions. Given the effects of online misinformation and widespread foreign involvement, this phenomenon also risks undermining democratic sovereignty through the unjustified inclusion of foreign actors in public deliberation.

- The harmful effects of misinformation are difficult to correct after exposure, making **preventative** measures that focus on reducing systemic and individual **vulnerabilities** a priority.

Digital Information Literacy

- Digital information literacy (DIL) is a **promising pathway** for empowering citizens and cultivating mass-level resilience to misinformation and harmful online practices - however, the slow pace of change and the scale of cognitive demands placed on citizens means that it should be treated as only one **part of a broader, multi-layered and multi-actor strategy** for tackling online harms.

- Efforts to improve DIL through school education in the UK face numerous challenges, the first of which is the **absence of a national strategic framework** for defining and measuring DIL. This results in confusion and impedes the development of priorities for action in this area.

- DIL-related skills and knowledge are **dispersed, disconnected** and **rarely applied** to the modern digital world under the **current curriculum**. The burden of realising the civic importance of DIL, making DIL-related connections across subjects, understanding how these apply to the digital information landscape, and subsequently practising this, lies almost entirely on students.

- There exist **no systemic DIL-related training, standards and support for teachers**, meaning that current provisions of DIL in schools are the province of individual teachers who recognise the importance of DIL education independent of curricular standards. As such, DIL provisions are likely to vary both within and between schools, and many teachers are likely to be under-prepared for future education reform on this front.

Key Recommendations

Tackling the risks from micro-targeting requires:

- **Improving oversight** through updating electoral legislation, introducing ad library standards, auditing machine-learning processes and restricting the pace and volume of communication to match the capacity of regulatory bodies.

- Considering how **incentives** for problematic action can be **altered** through stricter penalties and actions that introduce an element of broadcasting to micro-targeting — for example, by restricting political targeting to posts that have been published and archived on the official social media pages of political parties and candidates.

- **Empowering citizens** through more easily-accessible and thorough user controls relating to data use and targeting preferences, as well as digital information literacy efforts.

Combatting the risks from online misinformation requires:

- **Collaboration** and **sharing of responsibility** between government regulators, tech companies, media and influential public figures.

- Focusing on **preventative**, rather than corrective measures - including the collaborative delivery of 'inoculating' messages during critical periods.

- **Helping individuals to make the right choices** in the modern digital information landscape through the introduction of helpful nudges in platform design, mandatory training of political office-holders and mass-scale digital information literacy efforts.

- **Toughening of sanctions** to deter the spread of problematic content by those in positions of public visibility and responsibility.

Efforts to improve digital information literacy should prioritise:

- The creation of **common and comprehensive frameworks** for defining and assessing DIL, which can subsequently drive the development of a **national evidence base**.

- The introduction of DIL as **a core, cross-curricular component** at all levels of school education.

- Ensuring that DIL-related reforms are supported by **holistic transformations in teacher support and training**, as well as the **inclusive engagement** of all stakeholders in curricular reform.

Introduction

The arrival of the internet as a mass medium was initially hailed for its democratising potential. Digital technology became a tool for surmounting the offline realities of authoritarianism - democracy activists possessed a technological edge that could be used to circumvent physical restrictions on political rights and freedoms, avoid authoritarian monitoring, and spread information, techniques and ideas quickly across tightly-controlled geographic borders.

Today, the tables might be turning - digital technology is now thought to be creating new pathways for circumventing protections for political rights and freedoms, avoiding public scrutiny, and facilitating foreign interference. In particular, there are growing concerns that digital technology leaves citizens exposed and vulnerable to problematic political communication that can undermine the citizen consent, support and participation critical for the continued survival and functioning of democracy.

In the wake of the digital surge caused by the COVID-19 pandemic, citizens are spending a record proportion of their day on their digital devices,[1] whilst the level of public dissatisfaction with democracy is at its highest since the mid-1990s.[2] Thus it seems that in this twenty-first century, the *demos* have never been more online and democracy has never felt more at *risk*.

The present report seeks to advance our understanding of the challenges that the digital communication of political information poses for democracy and to help identify some priorities for action on this front.

The discussion summarises the major findings from a multidisciplinary literature review and three complementary academic workshops conducted by the Democracy@Risk Project at the University of Manchester in 2020. Specifically, it draws on existing research and debates in scholarship to assess two major sources of potential harm for democracy in the age of digital communication - **digital political micro-targeting** and **online misinformation**. In addition, it considers some key challenges relating to **digital information literacy**, as one of the most popular solutions advocated by observers and researchers.

Our first aim is to encourage deeper reflection about the origins and functioning of digital political micro-targeting and online misinformation, in the hope that improvements in understanding can lead to more comprehensive solutions that tackle some of the root causes, rather than the symptoms, of democratic malaise in the digital age.

Our second aim is to assess the evidence relating to the *effects* of these two phenomena and draw out some implications for democracy. Whilst there is a growing body of literature focusing on these two areas of potential harm, scholarship remains in its nascent stages, with few clear-cut answers generated in response to common concerns. The rapid pace of technological innovation and the great volume of potentially damaging digital political communication emitted on a daily basis, however, mean that we can scarcely afford to wait. As such, this report aims to piece together emerging insights from a range of disciplines and grey literature and ground them in conventional wisdom established through decades of research in political science and psychology, all with the purpose of generating some priorities for action in defence of democracy.

Finally, the discussion aims to propose some general recommendations for action, as well as to consider in greater detail the priorities for improving the provision of digital information literacy in UK schools, in the wake of growing calls for progress in this area.

Digital Political Micro-Targeting

Digital political micro-targeting is a form of online advertising that uses personal data about individuals to determine *whether*, *what* and *how* political adverts are shown to them on digital platforms such as social media, with the ultimate aim of influencing those individuals' attitudes and behaviour.

Originating in commercial advertising, this practice is becoming increasingly popular amongst political campaigners because it is thought to enable advertisers to reach key audiences with greater precision and efficiency. Yet there are also growing concerns over what this form of political communication means for electoral integrity, transparency, public discourse and policy.

This part of the report outlines the fundamental features of digital political micro-targeting and discusses the nature of the challenges it poses for democracy, based on our review of the contemporary debates and findings in scholarship.

How does it work?

As we browse and engage with content online, we leave behind a great volume of digital footprints containing potentially powerful insights into our lives.

Such footprints are generated whenever we share our personal information with organisations (for example, by opening an online account) or our online activity is recorded by tracking technologies, such as cookies. This information can be harvested and compiled into detailed personal profiles that list our habits, interests, lifestyle preferences and identity - some declared by us openly, and some assumed about us, based on the things we consume, like or search for online.

Digital political micro-targeting is a political advertising strategy that taps into this wealth of data to create a nuanced online campaign which tries to communicate and emphasise the right messages to the right audience.

As part of this practice, campaigns pay tech companies to feature a short political message to their users. Political adverts can subsequently appear alongside social media posts from our friends and family, as well as search engine results and content on the websites and apps we visit.

In doing so, campaigners are faced with the challenge of compressing complex, multidimensional political promises and arguments into a small advert that must persuade and stand out to individuals as they scroll quickly through content, often on the small screens of their smartphones. To overcome this problem, campaigners adopt a segmented approach to advertising, breaking down the audience and campaign arguments into smaller sub-groups.

Once political advertisers establish which segments of the audience are of key interest to the campaign, they turn to tech companies to find and reach this ideal audience among their users. Adverts might subsequently be targeted at specific individuals known to the campaign, as well as anonymous citizens, whose characteristics match an ideal profile. The personal data compiled by campaigns, third parties and tech companies is of paramount importance at this stage because it is used to determine whether individuals belong in the target group. The level of detail in these harvested profiles means that audience groups can be constructed on the basis of minute and specific criteria, which gives rise to the term '*micro*-targeting'.

Political adverts are subsequently shown only to those in the preferred audience categories. Campaigns can create different adverts for different audience groups on the basis of predictions about what motivates them, and test different versions of the same advert among sub-groups in real-time.

As a result, not every citizen receives adverts from a political campaign under digital political micro-targeting - and of those who do, not everyone necessarily sees the same type or indeed the same version of an advert.

How new is it?

Scholars disagree about whether digital political micro-targeting represents an entirely new phenomenon. Whilst some view this new practice as a product of the digital age, others have noted parallels with more traditional campaigning practices such as door-to-door canvassing, which are also used to connect with specific audiences.[3]

With this in mind, digital political micro-targeting emerges as a practice that descends from previous eras in political campaigning but utilises the latest developments in digital technology to enhance earlier capabilities to unprecedented levels. We note five major enhancements mentioned frequently by scholars:

1. More comprehensive voter profiling

Digital technology expands the breadth and depth of information a campaign has about the recipients of its adverts, making it possible to target individuals on the basis of more precise criteria beyond their *geographic* location or *demographic* characteristics (such as gender, occupation or religion).

With citizens spending increasing proportions of their daily lives on their digital devices, online activity produces a steady flow of detailed information about user interests, opinions and habits, which can deliver insights into the *psychological* attributes of voters. Different sets of vast personal data from different organisations are also often combined and matched to compile detailed multi-dimensional profiles on individuals.

2. Increased volume of tailored communication

As with all forms of digital communication, digital political adverts can be delivered in large numbers rapidly. Cheaper costs and huge audiences commanded by tech companies broaden considerably the potential reach of a campaign.

3. Automation of decision-making

Automated decision-making can play a significant role at various stages of digital political micro-targeting. Algorithms can be involved in the collection and processing of personal data, and are capable of making predictions about individual traits (such as personality) on the basis of user data. Tech companies also typically rely on algorithms for the delivery of targeted advertisements to their users, meaning that decisions over whether and which political adverts are shown to individuals are often the outcomes of machine-learning.

4. Monitoring campaign effectiveness in real-time

In the past, political organisations relied on surveys, dial tests and focus groups to test the effectiveness of political messages across audiences. Digital political micro-targeting expands further the opportunities for monitoring audience reactions. In addition to the analytical insights offered by tech companies, campaigns can use testing tools such as A/B testing, which presents variations of the same message to two groups in order to identify the most effective content. This data can be analysed in real-time.

5. More rapid adjustments to messages

Compared with more traditional forms of advertising, which cannot be withdrawn (as with postal communication) or can take a long time to update because they are dependent on the publication and broadcasting schedules of traditional media, digital political advertising is more flexible. The relatively small size of a digital advert and its limited content capacity means that new adverts can be created quickly. Adverts can also be approved by tech companies and begin reaching the desired audience within a matter of hours, if not minutes. Meanwhile, in response to real-time data about the effectiveness of different adverts, campaigns can cancel less effective versions quickly.

These enhancements should not be viewed in isolation from one another.

Traditional forms of political communication *can* deliver some of the above elements but only partially and separately. The biggest novelty of digital political micro-targeting therefore lies in its **cumulative** capabilities - the potential for utilising all five of the above enhancements at the same time.

FIGURE 1
Digital Political Micro-Targeting in numbers

43% of total advertising spending by political campaigns in the UK was spent on digital advertising in 2017 [4]

5.9 million Facebook ads run by the Trump campaign in 2016 [6]

$192.3 million spent by Trump and Biden campaigns on Facebook ads during Jan-Oct 2020 [5]

€17.3 million spent on **185,988** Google political ads in the EU and UK since March 2019 [7]

Common concerns

Common concerns about digital political micro-targeting can be divided into three broad types:

1. Persuasive capabilities: *Can digital micro-targeting 'steal' elections through subliminal persuasion?*

In the wake of the Cambridge Analytica scandal, concerns have been raised over whether political campaigns are becoming too 'scientific', with sensitive personal data leaving voters vulnerable to 'subconscious' forms of influence and manipulation. In particular, many question whether 'psychographic' targeting could be used to tap into hitherto-unavailable forms of subliminal influence, manipulate voter opinions and subsequently 'steal' elections.

2. Regulatory: *Does digital political micro-targeting receive adequate oversight?*

The second set of concerns questions whether digital political micro-targeting meets existing standards for political advertising which were created for traditional forms of communication, and focuses on the extent to which we are able to monitor this practice and check compliance. Many worry that this practice represents a 'wild west' in political advertising that falls short of the established standards for transparency and privacy and facilitates loopholes in electoral regulation.

3. Structural: *Does digital micro-targeting entrench problematic patterns in political behaviour?*

As digital political micro-targeting becomes more widespread, some scholars and observers are concerned that this could bring about a hyper-segmentation of the electorate that would transform political communication and public discourse, and incentivise political behaviour that could ultimately harm relationships between candidates and voters.

In the next section, we discuss the extent to which these concerns appear to be supported by findings in extant research and identify the resulting implications for democracy. Whist we find concerns over persuasive capabilities to be the least supported by research, regulatory and structural concerns have a stronger evidential basis, with significant challenges for democracy emerging on these fronts.

Evidence and implications for democracy

Can digital micro-targeting 'steal' elections through subliminal persuasion?

It is **unlikely** that elections can be 'stolen' through the subtle manipulation of voter opinions via digital political micro-targeting for the following reasons:

1. Micro-targeted political communication carries low persuasive power

Whilst personality has been linked to voter choices[8] and psychological targeting does appear to have some sway over consumer behaviour,[9] we find little empirical support for the claim that 'psychographic' micro-targeting is a powerful method of persuading voters to abandon or alter their pre-existing political beliefs and positions.

Scholars note that voters differ from consumers,[10] and thus tactics which are effective in commercial advertising might not be expected to achieve similar results in the political sphere. Specifically, most scholars agree that political beliefs are 'sticky' and resistant to change, making the persuasive effects of political advertising extremely small.[11]

Potential for subliminal manipulation is further limited by the fact that political messages delivered through digital micro-targeting are typically labelled and therefore recognised by voters as a form of advertising. This recognition is likely to reduce the persuasiveness of political communication since voters have been found to be suspicious of and hostile towards messages created and sponsored by political organisations, as opposed to the messages shared by their peers.[12]

2. Effects of digital micro-targeting are likely to be overestimated by practitioners

Scholars have noted that data consultants and campaign managers have multiple incentives to 'oversell' the precision and effectiveness of their tools.[13] The persuasive power of digital political micro-targeting is also likely to be overstated because the performance of an advert is largely measured in terms of *online* responses (such as clicks, sign-ups, or engagement). The translation of *digital* political behaviour into *real-life* behaviour, however, is difficult to measure and remains largely hypothesised.

In particular, researchers continue to disagree over whether online political activity is a *stepping stone* towards or a *substitute* for offline political participation, since citizens might overestimate the impact of online activism and stop short of real-life actions, having assumed that they have done their part.[14]

3. Micro-targeting is imprecise and risks backfiring

The effectiveness of micro-targeting techniques depends on the kinds of data available to campaigners and advertisers. This varies according the types of platforms selected for micro-targeting, as well as the national data protection and regulatory environment. For example, scholars have noted that the stricter regulations and different political traditions mean that, at present, many political parties in Europe cannot campaign like those in the US.[15]

Access to detailed data alone, however, does not guarantee accuracy in micro-targeting. For example, algorithms have been found to skew the distribution of micro-targeted ads in ways that are unintended by advertisers, meaning that audiences can differ greatly from the criteria specified by campaigns.[16]

Such impressions can *backfire* for candidates. For example, voters have been found to respond negatively to 'mistargeting' and punish political actors for receiving information meant for a different category of voter in error.[17]

4. Electoral impact of digital micro-targeting depends on the suitability of the overarching campaign strategy

In order for micro-targeting to serve as a powerful electoral instrument, political actors must first know why, when and how to use it. This requires them to be able to define their strategic goals accurately, devise accurate models of voter behaviour, identify the types of voters to target, and design a set of campaign promises which responds adequately to the needs of the target voter. Yet, political science has long held that political campaigns take place in the context of imperfect and limited information, which means that it is more appropriate to think of these processes as being grounded heavily in the political actors' *perceptions* of the political landscape, rather than objective truth.[18] This means that the overall electoral impact of micro-targeting depends on the accuracy of these initial perceptions and models, and the strategic calculations stemming from them.

A complex chain of influence

Overall, our review finds little evidence to suggest that digital micro-targeting is a precise tool with predictable and measurable effects on voter behaviour. Its effectiveness appears to depend on a complex (and poorly understood) set of conditions and vary according to the type of real-life action sought from the recipients.

Figure 2 visualises this complex relationship of factors is as a hypothetical 'chain of influence', which treats digital micro-targeting as one of many tools that could be used by campaigners to turn strategic goals (stage 1) into reality (stage 8).

Variations at each of these stages will affect the effectiveness of micro-targeting, and with it, the extent to which the use of this tool is decisive in terms of election outcomes. In order to meet such conditions and navigate the complexity of the legal, technological and political systems involved, political parties require extraordinary (and costly) professional expertise that blends together traditional campaign knowledge with data and computer sciences. Small amateur campaigns are therefore unlikely to reach this stage - however, scholars have highlighted a likely deficit in such expertise even amongst the more established experienced political parties.[19]

FIGURE 2
The chain of influence in digital political micro-targeting

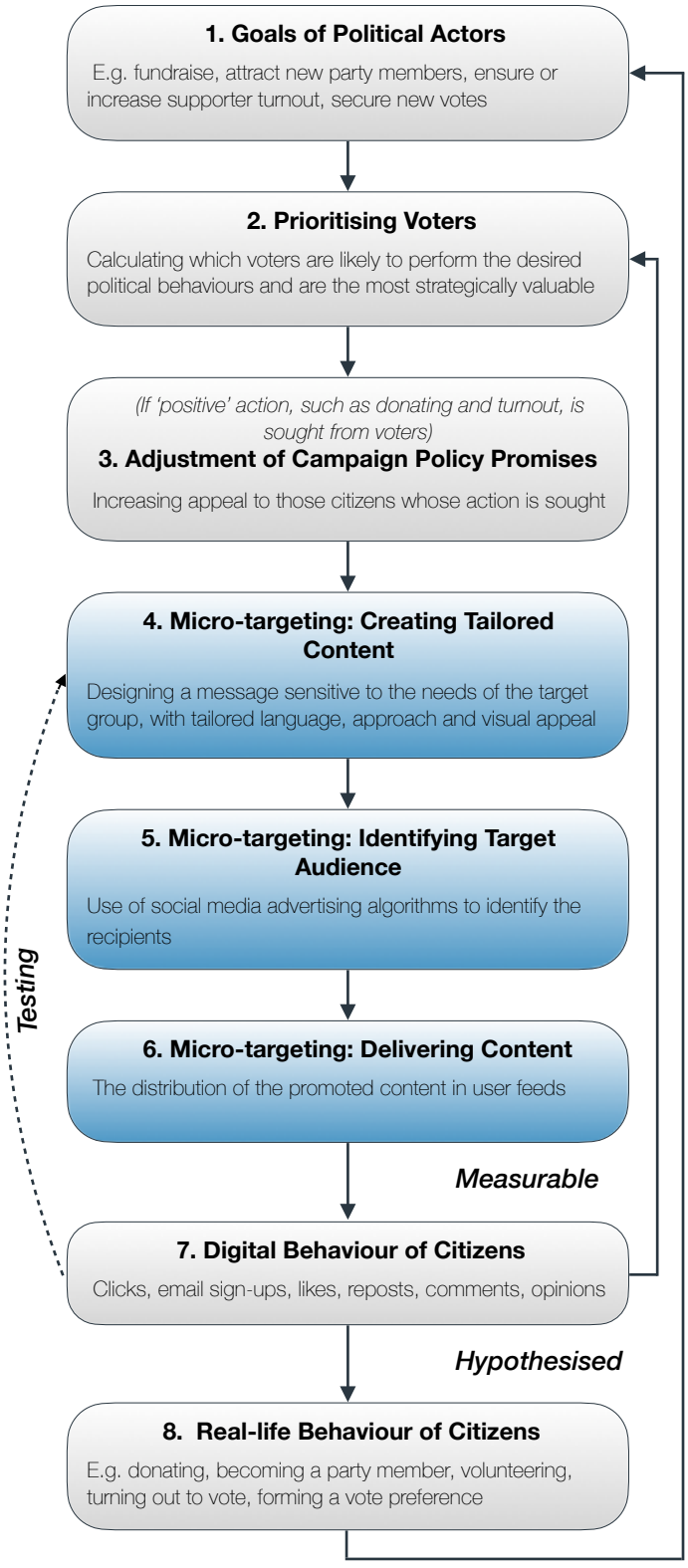


FIGURE 3

Gaps in oversight

Regulators

- UK electoral law last reformed in **2001**, before the era of digital campaigning and emergence of social media platforms [20]
- **No explicit requirements** for digital political micro-targeting under EU's General Data Protection Regulation (GDPR). Ambiguities in wording and derogations enable loopholes [21]
- UK political parties can process personal data on **'revealing political opinions'** without the need for voter consent, under the Data Protection Act 2018 [22]

Ad Monitoring

- **No** common standards for ad libraries and definitions of 'political' ad
- **Age, gender** and **location** are the only audience details shown in Facebook ad library
- Google **under-reported** spending on UK political adverts in 2019, in one case by a factor of 1,000 [23]
- **40%** of all UK political ads disappeared from the Facebook ad library two days before the 2019 General Election [24]
- Facebook algorithms have been found to **inhibit** a political campaign's ability to reach voters with diverse views and **skew** audiences in ways that are unintended by and unapparent to advertisers [25]

Citizen-level Oversight

- **Fewer than 15%** of data transfers to third-parties are disclosed in website privacy policies [26]
- It takes **31 minutes** to read Facebook's privacy policy [27]
- On average, privacy policies require **university-level** reading skills [28]
- **54%** of UK residents aware of digital political micro-targeting in 2019, with **44%** aware of the use of **'dark ads'** in digital micro-targeting (ads seen only by the recipient) [29]

Does digital political micro-targeting receive adequate oversight?

Digital political micro-targeting is plagued by **gaps in oversight and regulation**, some examples of which are highlighted in Figure 3. Our review identified the following challenges discussed frequently by researchers:

1. Shortfalls in regulatory provisions

- UK electoral legislation predates digital political micro-targeting and is out of touch with campaigning in the digital age

- EU's GDPR has been in operation for only three years, with the meaning of many rules, ambiguities and contradictions in derogations still to emerge through case law

2. Lack of ad transparency

- Ad libraries are not standardised across platforms and suffer from technical errors and inaccuracies[30]

- Ad libraries do not disclose the full set of reasons advertisers specify for targeting or the types and sources of personal data used to enable this

3. Opaque automated decision-making processes

- Facebook encourages advertisers to 'embrace a certain agnosticism towards placement, platforms and ... even audience'[31] by handing decisions over to machine-learning, yet the functioning of automated decision-making (such as algorithms, which can unintentionally skew the delivery of ads) is neither tested by regulators nor reported in detail

4. Inadequate transparency and agency provided to citizens as users

- The privacy policies set out by companies are too long, use complex terminology and do not include enough information about how data is shared with third parties, making informed consent unlikely[32]

- The ICO concluded that Facebook is not sufficiently transparent when it comes to enabling users to understand how and why they might be targeted by a political party or campaign[33]

Democratic Harms

In sum, there emerges a distinct lack of industry-wide standards with respect to record-keeping, transparency and user controls, meaning that at present, digital political micro-targeting is subject primarily to self-regulation by tech companies.

In the context of the huge volumes of digital political adverts distributed (Figure 1) and the widespread use of 'dark ads' (which can be seen only by the recipient), these gaps in oversight and control give cause for grave concern.

Specifically, two interlinked harms to democracy emerge as a likely consequence of such gaps in oversight.

Reduction of accountability

The obstacles to oversight and monitoring reduce the opportunity for public scrutiny, since a large proportion of political messages will not be seen by opponents, the media and the broader public. This in turn leaves open the possibility that political ads will contain false claims and contradicting promises that will not be challenged.

The ability to make narrow promises in the private setting of personal user feeds also risks misconstruing the overall political vision proposed by candidates and generating false expectations amongst voters. Without knowing and being able to prove the promises made by candidates, citizens are fundamentally limited in their ability to hold political actors to those promises during their time in office or to punish those actors for breaking promises during the next election - thus undermining the role of elections as an instrument of accountability.

Erosion of trust and democratic legitimacy

Widespread awareness of gaps in oversight risks diminishing public trust in political candidates and might undermine the perceived legitimacy of electoral outcomes. In addition, repeated negative experiences confirming a lack of accountability (as above) risk eroding public faith in democratic institutions.

Does digital micro-targeting entrench problematic patterns in political behaviour?

Digital micro-targeting is at base a tool for identifying an audience and using this insight to enhance the appeal of a political message as it reaches that audience. Its growing popularity therefore means that a greater proportion of political communication is becoming **selective** in terms of its audience and **tailored** in terms of appeal.

This increase in selective and tailored thinking in political campaigning is likely to encourage actions that harm democracy in the long-term, even if this consequence is not intended by the political parties and candidates themselves.

Consequences of selective communication

1. Political redlining
Determining that some voters are unlikely to turn out or to affect voting outcomes may lead campaigns to avoid investing time and money into contacting certain categories of voters. By helping campaigns to select and reach their preferred voters, digital micro-targeting therefore also helps them to *avoid* others.

The practice therefore enables a form of 'political redlining', where assumptions about political opinions of voters restrict the supply of political information to those citizens who are not perceived as pliable enough or valuable to a campaign.[34]

For example, a study of the Republican Party's voter files in Florida and Virginia found that young, less-educated, and lower-income registered voters were classified as unlikely to turnout at higher rates than older voters and those with higher levels of income and education, and were subsequently less likely to be contacted by the campaign.[35] Similarly, the Democratic Party's contact records also show that previous campaigns were less likely to contact those voters whose propensity to turn out and support the party was estimated to be low.[36]

2. Voter Suppression
Knowledge of individuals' unfavourable voting intentions can be applied not only passively (to avoid them) but also actively (to target them).

Whilst micro-targeted political communication is predicted to have very low likelihood of persuading voters to shift their political allegiances, it may nevertheless deliver strategic gains if it is able to turn voters away from *participating* at election time. Digital political micro-targeting can therefore be used for voter suppression - attempts to discourage individuals from turning out to vote on election day.

The precedent for this has already been established in recent elections - for example, the Trump campaign acknowledged its use of three voter-suppression campaigns, which sought to discourage the turnout of African-Americans, young women and 'idealistic white liberals' by targeting those voters with messages highlighting Hillary Clinton's comments about African-American men in the 1990s, Bill Clinton's behaviour towards women, and Hillary Clinton's support for the Trans-Pacific Partnership deal, respectively.[37]

Democratic Harms

Reduced participation and inclusion
Voters are more likely to participate at election time when they are contacted by political campaigns,[38] which means that ignoring or excluding certain categories of citizens from such campaigns will likely fuel voter disengagement and entrench low voter turnout amongst those groups. In this sense, the candidates' perceptions of the electorate may become a self-fulfilling prophecy.

Whilst voter suppression may be electorally advantageous to campaigns in the short-term, ultimately, this tactic is dangerous because it encourages voters to express their preferences by switching off from *democracy*.

Together, these unequal promotions and protections of electoral participation undermine the principle of empowered inclusion necessary for democracies to function and maintain legitimacy.

Consequences of tailored communication

1. Particularistic policy promises

Greater reliance on digital micro-targeting puts political parties and candidates at risk of pursuing policy with particularistic benefits, instead of delivering broader public goods.

Scholars have modelled the behaviour of political actors to show that when candidates are able to target messages to specific groups of voters, the result is greater commitment to projects that benefit small groups, even if the social cost of such projects outweighs the benefits.[39] Digital political micro-targeting therefore risks resulting in inefficient policy platforms which do not represent or respond to the needs of the electorate at large.

2. Overemphasis on wedge issues

It is highly likely that greater reliance on digital political micro-targeting will result in the dominance of divisive issues at election time.

An earlier study of traditional 'offline' campaigning in the USA has shown political actors are more likely to focus on 'wedge' issues (divisive topics such as immigration, crime or sexuality which cause conflict within an otherwise united group) in targeted forms of communication, when compared with messages which are more widely broadcast (such as television advertising).[40] Scholars have also constructed models which show that the ability to target messages is likely to result in more extreme positions being taken on wedge issues by political candidates.[41]

Democratic Harms

Weaker representation and collective agenda

The overwhelming focus on wedge issues is not only divisive but also raises considerable problems for representation. Voters may be misled towards supporting those candidates who, whilst representative in terms of their position on a wedge issue, nevertheless represent only a sliver of the voter's beliefs and political preferences. The narrow focus of micro-targeted adverts makes it difficult for voters to judge the extent to which a candidate prioritises an issue relative to other promises in the full version of their manifesto.

Continued emphasis on wedge issues in micro-targeted messages may also overstate the salience of this issue relative to other social problems.

Taken together, these features of digital political micro-targeting make it likely that voters will not make the optimal choice in terms of representation from the candidate options available to them, which will likely fuel voter disillusionment with politics.

In addition, the fragmentation of policy promises through digital micro-targeting makes it more difficult to draw conclusions about collective will and may therefore impede the delivery of a clear governing mandate for candidates.

Recommendations

Strengthen Oversight

Digital political micro-targeting must be made more transparent to enable accountability and maintain public trust. Common industry standards are needed and the balance of regulation must be moved away from tech company self-regulation towards independent monitoring bodies.

Actions can include:

- **Updating UK electoral legislation** to establish specific rules and reporting processes for digital campaigning and spending
- Creating **industry standards for ad libraries**. Records should include full details of the targeting, optimisation and ad placement conditions specified by advertisers, and more granular data on spending, impressions and type of political advertiser [42]
- Regular **audits of ad delivery algorithms** for skews in distribution and possible misalignment with data protection regulation [43]
- **Limiting the number of ads** a political campaign can run per week to help over-stretched regulators keep up with monitoring ad content in real-time

Alter Incentives

The ability to distribute more selective and tailored messages encourages political actors to engage in actions that are legal but nevertheless damaging for democracy in the long-term. Improvements in oversight alone will not be able to prevent this problematic political behaviour.

We recommend focusing on deterrence measures which reduce the appeal and raise the costs of problematic political behaviour.

Actions can include:

- Introducing **hybridity** to digital micro-targeting. For example, introducing broadcasting to micro-targeting by allowing the targeting only of those posts which are published (and permanently archived) on the official social media pages of political parties or candidates.

- Forcing campaigns to issue **notices** or **apologies** to those users who have been exposed to ads that breached regulatory standards

- Raising the maximum **financial penalties** that can be issued by regulators [44]

Empower Citizens

Empowering citizens can make oversight more multidimensional and build critical resilience that reduces the harmful effects of digital micro-targeting in real-time, particularly if problematic political ads manage to escape the regulatory filter.

Actions can include:

- Enabling users to **opt-out** of political advertising online
- Providing users with **easily-accessible** and **non-binary control tools** over the kind of data that can be used to track and target them
- Investing into **digital information literacy** programmes for citizens

Online Misinformation

As digital technology lowers the barriers for entry into political analysis and reporting, and the dissemination of information is made easier, cheaper and faster with social media, concerns emerge over a wide range of information 'disorders' which plague the contemporary digital information landscape.

Whilst collectively, these phenomena are often referred to under the umbrella term of 'online misinformation', the following sub-types are frequently discussed:

Misinformation
False or inaccurate information spread without the intent to cause harm.
Disinformation
Information created and disseminated with the intent to deceive or mislead
Malinformation
The dissemination of true facts obtained through illegal means (such as leaks and hacking) to inflict harm on others and achieve strategic goals.

Our review finds that the distinctions between the above categories are more **blurry** than they might seem initially. As discussed later in this section, online information exists within a complex eco-system comprising many different actors, outlets for publication and pathways for transmission. As it moves through this system, problematic content may take on different forms - for example, misinformation can be used to spark disinformation operations, whilst disinformation claims can become misinformation when they are picked up by well-meaning but misled actors.[45] Distinctions between the above sub-categories are further complicated by the difficulties of establishing the origins of problematic information and attributing intent, as well as the interchangeable use of these terms in studies.

As a result, in this report, we refer to 'online misinformation' broadly, as an umbrella term that covers all three sub-types outlined above.

Beyond 'fake news'

Whilst the notion of 'fake news' has gained traction in public discourse, online misinformation is more complex than this term suggests. Misinformation is becoming more difficult to spot both because of the increasingly high quality of fake content (as with 'deep fake' software that can swap and manipulate faces and voices) and because it is not limited to outright falsehoods but can include more sophisticated narratives constructed on the basis of selected, contested and/or misinterpreted claims.

For example, early in the pandemic, one study of misinformation around COVID-19 found that the most prevalent forms of misinformation (representing over 59% of the problematic content and 87% of the social media interactions investigated) were not objectively false claims, but represented existing and largely accurate information which had been twisted and repurposed in ways that made it false or misleading.[46] In addition, misinformation is not always aiming to boost beliefs in false claims - problematic techniques include 'fact-softening', which involves doubting facts and the subsequent undermining of beliefs in the veracity of factually accurate information.[47]

Evidence and implications for democracy

Who creates online misinformation?

State actors
Intelligence services and state-sponsored agencies in Russia,[48] Iran[49] and China[50] have been found to engage in extensive digital information operations on social media with the aim of influencing public discourse and politics in other states. State actors have also been found to conduct disinformation operations targeted at domestic politics, as with the South Korean National Intelligence Service during the nation's 2012 presidential elections.[51]

Media
Media outlets such as the Russian state-sponsored RT and Sputnik are known to publish information that has been twisted into misleading 'strategic narratives'.[52] Research has also uncovered the publication of inaccurate and misleading information (and the subsequent concealment of this) by the mainstream media in the West, particularly during the early stages of rapidly-developing crises and shock events.[53]

Third-party agents
Media reports show that extensive disinformation operations have originated from teenagers in Macedonia,[54] individuals in Romania,[55] and commercial actors in the USA.[56] US conspiracy theorists have also been identified as the original creators of problematic digital information, even when such information is subsequently picked up by foreign state agents.[57]

Public figures and Influencers
Individuals with large online followings and public figures, including celebrities and political office-holders, can generate large volumes of problematic content through their personal platforms such as blogs and social media accounts (Figure 4). US President Donald Trump, whose Twitter content has carried the platform's misinformation warning labels on multiple occasions, remains perhaps the most high-profile example of this.

Citizens
Digital technology has reduced the barriers for entry into content creation and publication, resulting in the democratisation of the modern information landscape. Individual citizens can create false or misleading information through social media posts featuring personal reports, opinions and video/photographic content.

FIGURE 4
The 'Disinformation Dozen' [58]

12 individuals identified as **extremely influential** creators of anti-COVID-19 vaccine content online

65% of **anti-vaccine content** on Facebook and Twitter between 1st February and 16th March 2021 was attributable to this 'Disinformation Dozen'

Why?

Misinformation is born from a wide range of motives, including:

- Electoral influence
- Financial gain
- Destabilisation for geo-political gain
- Damaging the credibility of opponents
- Humour
- Genuine error (no ill intent)

How does misinformation grow?

There are many pathways through which problematic content could be spread further, or *amplify*.

Crucially, misinformation can be spread not only by those who created or believe in it but also by those attempting to stop it. Thus, in seeking to warn others of misleading and potentially harmful content, individuals and organisations may be inadvertently contributing to the amplification of misinformation.

Co-ordinated inauthentic amplification
'Fake' actors, such as 'bots' and accounts under false identifies, can be used to disseminate misinformation and create an impression of widespread support.

For example, research into disinformation operations has uncovered complex disinformation armies.[59] Typically, these take on a hierarchical structure, at the top of which a handful of human social media accounts are seeking to gain momentum in public discourse and integrate into organic communities online. Underpinning these are thousands of 'fake' automated social media accounts which typically project false communities and mimic grass-roots support - the latter practice is often referred to as *astro-turfing.*[60]

FIGURE 5
Examples of inauthentic behaviour on social media

5.8 billion fake accounts shut by Facebook in 2020 [61]

150,000 false accounts amplifying a co-ordinated influence campaign run by the People's Republic of China were shut by Twitter in June 2020 [62]

Organic amplification
Problematic content does not rely exclusively on fake actors or its creators to spread. Traditional media, as well as influencers and organisations, are known to advance the spread of misinformation online.

FIGURE 6
Challenges for journalism

80% of journalists admitted to being **tricked** by false information at some point in their career [63]

15% of journalists have taken part in any **training** on how to combat the effects of false information [64]

FIGURE 7
The role of influencers in the spread of online misinformation on COVID-19 [65]

Public figures are the most **influential spreaders** of misinformation about COVID-19

69% of total **engagement** with online misinformation was found to have been generated by posts published by politicians, celebrities and other public figures

In addition, research is beginning to highlight the role of citizens as curators and amplifiers of problematic information.[66] In the digital age, citizens are emerging as gatekeepers of information, who, as social media users, are choosing news not only for their own consumption but for that of others. This means that published information depends on 'user-generated visibility' in order to be impactful today.[67]

Online misinformation as an eco-system

The wide range of actors, outlets and interactions involved in online misinformation means that this phenomenon can be better understood as a complex eco-system, characterised by cycles of content creation and amplification, and a metamorphosis of facts and falsehoods within shared narratives.

Within this, there exists no unified point of entry, which makes tracing and stopping misinformation as it spreads immensely challenging. The eco-system makes online misinformation akin to a '360 degree' experience[68] for citizens, which means that a multidimensional approach is required to weaken some of the connections in this system and carve out gaps in the fabric of online misinformation.

Is online misinformation harmful for democracy?

Effects of misinformation

Reinforcement of political beliefs
Research suggests that misinformation carries a stronger *affirmative,* rather than *persuasive,* effect. Studies have shown that users are more likely to share misinformation when it aligns with pre-existing political beliefs.[69] More broadly, individuals are thought to be more likely to process political information on the basis of *directional motivation* - the desire to find support for conclusions which align with one's pre-existing beliefs - which makes citizens less critical of information which is consistent with one's attitudes, and more sceptical of information which runs contrary to one's preconceptions.[70]

Erosion of trust and growth of cynicism
Research suggests that misinformation risks an almost irrevocable erosion of public trust in political institutions and actors. For example, a recent study has found that online rumours in China decrease citizens' trust in government, which can be hard to recover for political actors since it requires the provision of irrefutable and vivid evidence that is often hard to obtain.[71]

In addition, growing misinformation and contestation of facts can foster the belief that factual claims, including those made by experts, are always subject to politically-motivated interpretation - leading to widespread cynicism as a default in political communication.[72]

Apathy, resignation and withdrawal

Widespread misinformation - or the perception of such, as a result of cynicism - increases the cognitive demands for citizens, which can feel overwhelming and breed paralysing uncertainty. A recent survey found that many US citizens are reducing their overall intake of news in response to concerns over the preponderance of 'fake news', with less politically aware individuals doing so at particularly high rates (50%).[73]

Polarization

Deceptive inauthentic behaviour online can fuel polarization, reducing understanding and respect between groups. As suggested by a recent study of information operations relating to the #BlackLivesMatter discourse, fake social media accounts can impersonate individuals from specific backgrounds and members of specific social groups, engage in divisive rhetoric and spread of misinformation, and subsequently both pull like-minded individuals closer whilst pushing opposing sides further apart.[74]

The affirmative effects of misinformation can also fuel gaps in political assessments and attitudes. A recent US study, for example, found that misinformation reduces political trust amongst opponents of the party in government, whilst increasing trust amongst government supporters, which suggests that the affirmative effects of misinformation could widen the rift between voters with opposing political views.[75]

Democratic harms

These effects of misinformation carry the following implications for democracy:

- Entrenchment of harmful political attitudes that are not supported by evidence
- Loss of faith in political actors and institutions, which undermines democratic legitimacy
- Reduced impact of government guidance, which is likely to be distrusted
- Reduced levels of political participation
- Difficulty with the formation and acceptance of collective will and agenda due to increased polarisation

In addition, the facilitation of these effects via inauthentic behaviour co-ordinated by foreign state and third-party actors means that actors that would normally lack the legal rights to participate in democracies are able to gain entry into the arenas of public discourse and participate in public deliberation through deception, thus undermining the principle of democratic sovereignty.

Can misinformation be corrected?

Whilst fact-checking practices are valuable, they are **not enough** to protect voters from the influence of misinformation.

On the one hand, corrections have been found to reduce misperceptions and the persuasive power of fake news stories, thus resulting in more accurate beliefs.[76] Yet important caveats remain:

1. Fact-checking organisations have limited reach and resources
The limited financial and labour resources of fact-checking organisations in the face of a vast eco-system of misinformation mean that fact-checkers have to prioritise the investigation of some claims over others and are constrained in the dissemination of their findings. As a result, many problematic claims remain untested by fact-checkers and many citizens remain oblivious to factual corrections.

2. Corrections can be rejected by citizens and backfire
Fact-checking corrections appear to be ineffective in the face of strong pre-existing political attachments and motivations. For example, individuals have been found to be more resistant to factual corrections when they share a political affiliation with those accused of making false statements.[77]

Studies are also beginning to highlight some possible variation in the effectiveness of fact-checks amongst men and women,[78] as well as across personality characteristics, such as the tolerance towards negativity.[79]

Moreover, researchers have reported instances of a 'backfire' effect, where factual corrections unintentionally *reinforced* false beliefs amongst the politically-knowledgeable supporters of those targeted by fact-checks.[80]

3. Changes in citizen attitudes and behaviour are not guaranteed even if corrections are accepted

Studies have shown that even when people acknowledge information as incorrect as a result of fact-checking, their sentiments towards political actors remain largely unchanged.[81] This suggests that even when fact-checking is successful in its immediate goal of correcting misconceptions, the *political* impact of these corrections is likely to be minimal.

Moreover, research in psychology suggests that misinformation continues to exert potentially harmful influence on citizen behaviour long after it has been 'successfully' corrected. Scholars have long emphasised a phenomenon known as the *continued influence effect*, which sees individuals retaining the conclusions reached on the basis of inaccurate evidence, even when such evidence is subsequently shown to be false.[82] More recently, one study has shown that misinformation results in political 'belief echoes' - effects on political attitudes which linger after misinformation has been effectively corrected.[83]

An imperfect filter

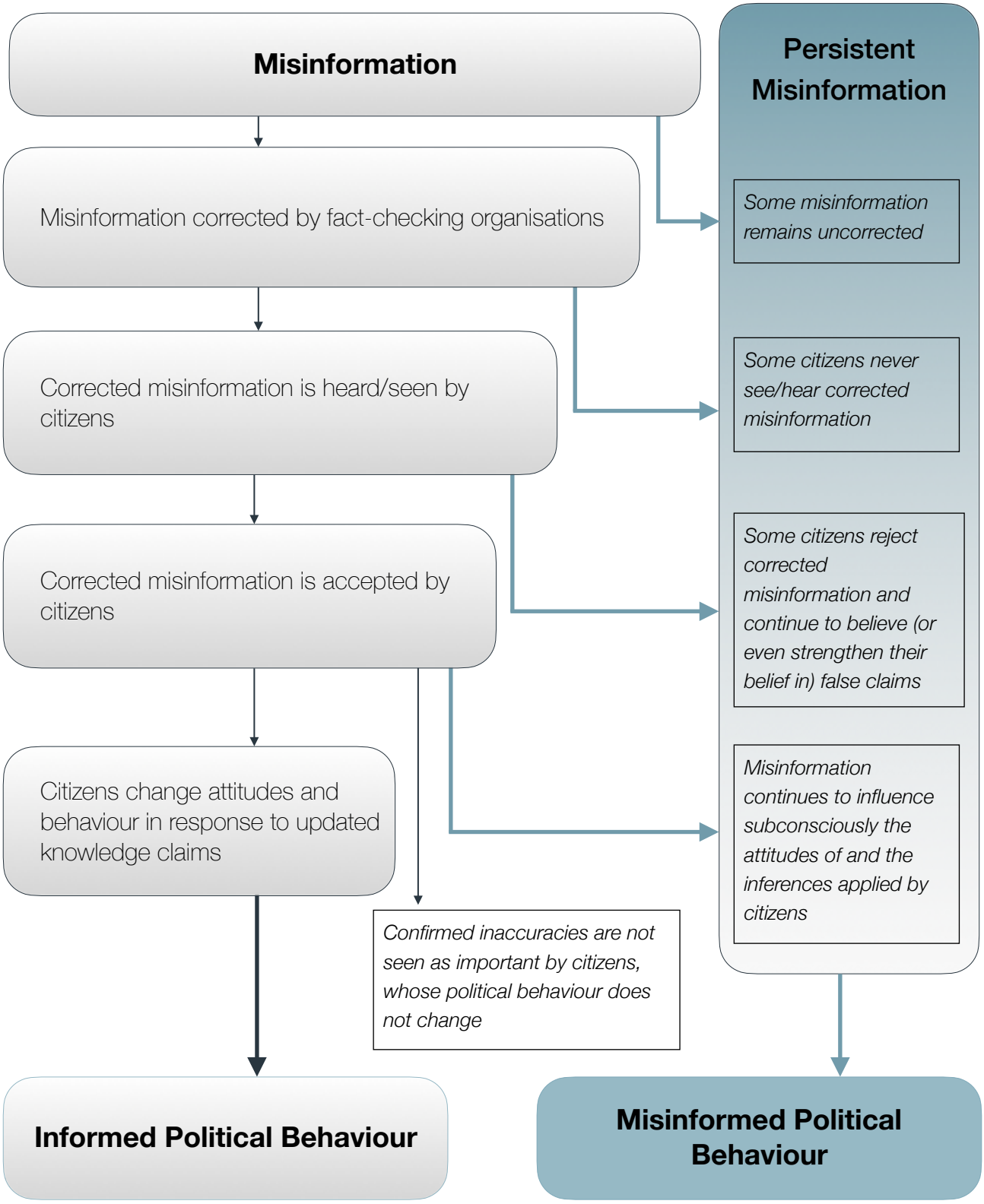
Overall, our review concludes that fact-checking can be thought of as a misinformation filter, which has the potential to neutralise some of the harmful effects of misinformation and facilitate informed political choices. Figure 8 visualises this concept.

Some misinformation manages to escape the filter at each stage of this process, however. Thus, false claims which have been effectively neutralised by fact-checking organisations represent only a small proportion of both the total volume of misinformation and the total claims corrected by the organisations themselves.

Fact-checking is therefore a potentially useful tool in the fight against misinformation, but its effectiveness is far from guaranteed - indeed, in certain cases, this tool can become counterproductive.

As such, we conclude that fact-checking should be seen as a last resort, rather than the first line of defence in the fight against misinformation.

FIGURE 8
Fact-checking as a filter of misinformation



Can we 'inoculate' against online misinformation?

Drawing on a biological metaphor, the 'inoculation' approach holds that, just as vaccinations protect people from a possible viral infection in the future, so can preemptive messages make people more resistant to future attempts at persuasion.[84]

Inoculation messages consist of two key elements - *threat* and *refutational preemption*.[85] First, such messages seek to expose a potential threat and warn individuals that they are vulnerable to persuasive attacks, which is thought to motivate the recipients to act defensively. The refutational element of inoculation subsequently seeks to equip individuals with arguments and evidence that can be used against a possible future attack and serves as an opportunity to practise counterarguments.

Potential

Overall, there is considerable evidence to suggest that inoculation is an effective method of building citizen resistance to misinformation. The results of a meta-analysis in 2010 have shown that, in a sample of 41 studies involving over 10,000 participants, individuals who were exposed to inoculation treatments were found to be more immune to information attacks than those who were not inoculated and those who received information that simply tried to bolster existing attitudes prior to an attack.[86] Crucially, inoculation exercises have been found to be effective even amongst individuals who have already been exposed to misinformation.[87]

Of particular interest is the potential for inoculating against arguments that are different from those included in the original refutation element of the inoculation, since this offers greater flexibility in the face of the rapid pace of technological and political developments. Recent studies have begun to indicate the viability of a 'broad-spectrum vaccine' against misinformation, which allows for the immunisation of individuals against misinformation relating to a wider range of subjects, by exposing participants to weakened examples of the techniques used by misinformation actors and by allowing participants to actively construct their own counter-arguments (or 'anti-bodies').[88]

The positive results from such studies give some reason for optimism, since they suggest that it is possible to build resistance to misinformation in a relatively short space of time.

Limitations

We find that inoculation messages should not be seen as a shortcut to raising the levels of digital information literacy within the population at large, however.

At present, it is difficult to envisage how inoculations can become mass-scale - in this sense, there exists a big difference between biological vaccinations and informational inoculation.

There are also emerging concerns about the extent to which the effects identified by the inoculation literature are long-lasting and are replicable in real-life, beyond the experimental settings.

Moreover, even if such barriers are overcome, there are concerns about the extent to which inoculations will be able to remain effective treatments against misinformation once such practices become public knowledge. For example, one study has shown that *metainoculation* (messages which describe how the inoculation processes work and encourage participants to think for themselves) can reduce the effectiveness of subsequent inoculation messages, which makes this tactic vulnerable to exploitation by malign information actors.[89]

Recommendations

Collaboration and sharing of responsibility to prevent and deter the spread of misinformation

We recommend a multi-dimensional response which should:

- recognise the **involvement**, **interdependence** and **responsibility** of multiple actors in the eco-system of misinformation
- help these actors to make the **right choices** in the context of an ever-evolving digital information landscape
- prioritise **preventative** over corrective measures, and
- impose appropriate **sanctions** for non-compliance or negligence

Actions might include:

Government

- Collaboration with tech companies to establish **practical industry standards** for monitoring, reporting, labelling and removing problematic content and its creators

- Using insights generated by government anti-misinformation task forces to **anticipate** growth in specific misinformation claims and develop '**inoculation**' programmes in response

- Providing **mandatory training** on digital information literacy and the responsible publication of content online for all **political office-holders** and their staff.

- **Raising the financial penalties** for misinformation in political ads online in proportion with contemporary levels of campaign spending

- Developing a comprehensive strategy for assessing and improving the levels of **digital information literacy** among citizens, including its introduction into the national curriculum

Tech Companies

- Collaboration with other tech companies to **share analytics** on known and suspected disinformation operations

- **Toughening sanctions** for repeat offenders, including lower thresholds (such as a small number of strikes) and a wider range of penalties (such as temporary bans from certain platform features, restriction of ability to post content without review, or restricting the visibility of posts to current followers)

- Attaching **preventative**, rather than corrective warning **labels** to problematic content

- Collaborating with governments and NGOs to distribute '**inoculating**' messages during critical periods to users free of charge

- Encouraging users to **think twice** and **fact-check before** posting or sharing links to political content

Media

- Newsrooms equipping journalists with professional **content verification tools**

- Offering digital information literacy **training** for journalists in all specialisms, not only to those reporting on disinformation

- Establishing **guidelines** for reporting on misinformation to minimise inadvertent exposure to false claims and the risk of fact-checks 'backfiring'

In Focus: Digital Information Literacy

The recommendations outlined in the previous parts of this report touch on the common theme of digital information literacy (DIL). In the fight against online misinformation, DIL represents a *preventative* measure that is likely to be key for building mass resilience in the long-term. With respect to digital political micro-targeting, DIL becomes relevant as a way of *empowering* citizens to oversee political communication targeted at them and recognise its effects.

This common theme is echoed in the recommendations of many researchers and observers, who have emphasised the importance of boosting the digital information literacy provisions within the national school curriculum as one of the key ways of combatting misinformation and political manipulation. For example, the UK Digital, Culture, Media and Sports (DCMS) Committee has argued that digital literacy should be treated as the 'fourth pillar' of education, alongside the traditional core components of reading, writing and maths.[90] More recently, similar recommendations for the embedding of 'critical digital media literacy' throughout the national curriculum were put forward by the House of Lords Select Committee on Democracy and Digital Technologies.[91]

In this final part of our report, we explore the challenges relating to digital information literacy as a solution to the potential harms generated by digital political micro-targeting and online misinformation.

What is Digital Information Literacy?

There exists no overarching consensus on how to define DIL, or, indeed, whether one should be using the term 'digital information literacy' in the first place. Whilst different scholars, organisations and institutions may insist on specific and unique meanings for the terms relating to DIL, in practice, different definitions frequently overlap and the same definitions are not applied consistently.

Having identified key themes within the traditions of 'media', 'information', 'critical-thinking' and 'digital' literacy studies, we have consolidated them into a more comprehensive, blended definition of digital information literacy, below.

Definition

We understand digital information literacy as the active use of the functional, technological and critical-thinking skills and knowledge, as well as the attitudes, necessary for locating, understanding, evaluating and creating information effectively and ethically in the ever-changing digital information landscape.

This definition entails five key dimensions of DIL:

1. **Functional digital skills** for the efficient navigation and use of digital technologies

2. **Technical understanding** of the digital information landscape and **how** information is created, published, stored, found and distributed through digital technologies

3. **Critical and interpretive skills** for the effective appreciation, verification and interrogation of information

4. Effective **simultaneous practical application of the above skills and knowledge**, and willingness to engage in meaningful **behavioural change**

5. **Communal sense-making** which seeks to build shared meaning, and acknowledges the shared responsibility of individuals, organisations and institutions

Characteristics of DIL as a solution

Our multi-layered definition reveals much about DIL as a solution to the challenges posed by digital political micro-targeting and online misinformation.

First, it highlights the cultivation of DIL as a complex, laborious and time-consuming task. For this reason, school education is often seen as the primary gateway for DIL - schools facilitate the early intervention, the delivery and consolidation of knowledge, and the prolonged supervised practice of key skills necessary for instilling effective and long-lasting DIL habits amongst the majority of citizens. As such, improvements in mass-level DIL levels are likely to be generational and slow-paced.

Second, at its core, DIL concerns the enhancement of capabilities relating to the use and creation of *information* in the digital age. Its focus on making sense of the world around us means that DIL is more closely linked to academic activity in schools than it is to the issues of online safety or personal development.

Third, adopting a more comprehensive definition of DIL also serves to recognise the full scale of demands that citizens face on a daily basis, if they are to operate effectively in the modern digital information landscape.

These characteristics of DIL mean that recommendations for greater investment in DIL must be accompanied by the following caveats:

- DIL should be treated as one part of a broader strategy that shares the burden of oversight between many different types of actors and does not place all responsibility on citizens
- The slow realisation of DIL efforts through education should be accompanied by smaller, short-term interventions that help citizens who are no longer in school education to make better choices within the digital information landscape
- DIL should be seen as separate from, albeit complementary to, online safety and personal development

Effectiveness of DIL in education

Education-based interventions appear to be effective at arming individuals against problematic information. One recent study has found that media literacy improved the ability of students to judge whether a post contained accurate information, even when posts containing misinformation were aligned with their pre-existing political perspectives.[92] Recent evidence also suggests that students can be trained to think critically about, and detect, problematic claims from a very young age - for example, primary school pupils in Uganda were found to be able to make more informed health choices after they were taught 12 simple concepts essential to assessing claims about health treatment.[93]

Finland, which has been recently ranked as Europe's leading country in terms of preparedness to withstand the impact of misinformation,[94] is setting a promising example for building mass-scale resilience to problematic political content in the digital age through the integration of DIL into the national curriculum. Specifically, all levels of the Finnish school curricula were redesigned between 2014-2017 to focus on the development of 'twenty-first century skills', which included the introduction of 'multiliteracy', digital competence and media education as core curricular components covered by all subjects.[95] This skills-based approach is underpinned by the mandatory delivery of at least one multidisciplinary learning module per year to encourage the practice of making complex connections between subjects and skills conducive to DIL.

DIL education in the UK

Current DIL provisions in the UK are facing the following challenges:

1. Absence of an overarching strategic framework for the provision of DIL through primary, secondary and further education
There exists considerable disagreement and confusion over what 'digital information literacy' entails, or, indeed, what it should be called: 'media literacy', 'digital literacy', 'critical literacy', 'information literacy', or combinations of these terms have all been used to refer to the same set of broad ideas by policy-makers and researchers. Moreover, there exists no specific framework that breaks down and sets a standard for DIL-related skills and knowledge, meaning that no coherent plan has been outlined to assess citizen preparedness to withstand the negative impacts of digital political micro-targeting and online misinformation.

Our review has struggled to identify comprehensive, multi-dimensional indicators of DIL in UK schools. The partial insights that do exist, however, paint a bleak picture (see Figure 9).

FIGURE 9
Tackling 'fake news' in schools [96]

2% of **children and young people** in the UK have the **critical literacy skills** needed to tell if a news story is real or fake

53% of **teachers** believe that the national curriculum **does not equip** pupils with the skills to identify fake news

The absence of a coherent definition and framework for the measurement of DIL also sows confusion over policy recommendations, risking important suggestions being overlooked. For example, conflicts have emerged between the conclusions reached by the DCMS Committee and the House of Lords Select Committee on Democracy and Digital Technologies on the one hand and the UK Government on the other. The former have both noted inadequate DIL provisions and called for curricular transformation (as discussed earlier), whilst the latter holds that the development of DIL is already enabled by the school curriculum.[97]

Improvements in conceptual understanding will help to prevent miscommunication between different governmental branches, departments and regulators. Greater clarification on this front will also help to direct and narrow future recommendations for policy - for example, instead of advocating general improvements in DIL, researchers can point to more specific gaps in curricula and priorities.

2. Opportunities for DIL within the school curriculum do not translate into action
Whilst elements of DIL are present within existing curricular frameworks, they are separated and dispersed across subjects[98] - meaning that students rarely get explicit guidance on and the chance to practise civic online reasoning.

For example, whilst traditional subjects like English, History and Science foster critical thinking and respect for evidence, these skills remain largely offline and rarely explain the value of the internet as a fact-checking tool, as well as specific strategies for differentiating bias from disinformation online. Similarly, the Citizenship curriculum is concerned with knowledge relating to the political system and the press, yet largely ignores the digital information environment and online civic skills. By contrast, Computing subjects boost the students' functional digital skills, but do not connect these to the evaluation of online information or explain how technological developments (such as search engine and social media algorithms) can facilitate misinformation and manipulation.

In sum, it seems that these three branches of the school curriculum represent important opportunities for the learning and application of digital information literacy skills amongst young people. Yet, at present, instead of mutually supporting and reinforcing key messages and practices that may help young people of today to become the informed citizens of tomorrow, the school curriculum diffuses digital information literacy.

The current curriculum therefore places the burden of connecting subject-specific knowledge and skills, translating them into DIL capabilities and learning to apply these to the digital information landscape entirely on the students.

Figure 10 visualises the above disconnect. To improve provisions for DIL, the national curriculum does not necessarily need to prioritise the introduction of DIL as a separate and new subject - rather, such literacy requires **further assembly** through the practice of online reasoning in and the fostering of collaboration between existing school subjects.

3. DIL provisions are not underpinned by large-scale systematic support for teachers

Recent findings show that teachers are important gatekeepers in educational innovation, with one study finding that computing teachers are able to reject innovations in digital technology curricula and hinder educational reform in practice, even when this carries a legal mandate and support from industry.[99] As a result, teacher-level factors, including teacher perceptions of the importance of DIL in education and the extent to which they are prepared to incorporate it into their teaching and learning, must be taken into consideration in discussions of DIL education in the UK. Given the multidisciplinary demands of DIL, such considerations should not be limited to computing teachers but instead account for teachers from a wide variety of subject backgrounds.

In the absence of a national DIL framework and standard for assessment, however, we know relatively little about how well teachers are equipped for the delivery of DIL-related skills and knowledge or prepared for educational reform on this front. The partial insights we do possess, however, give cause for concern.

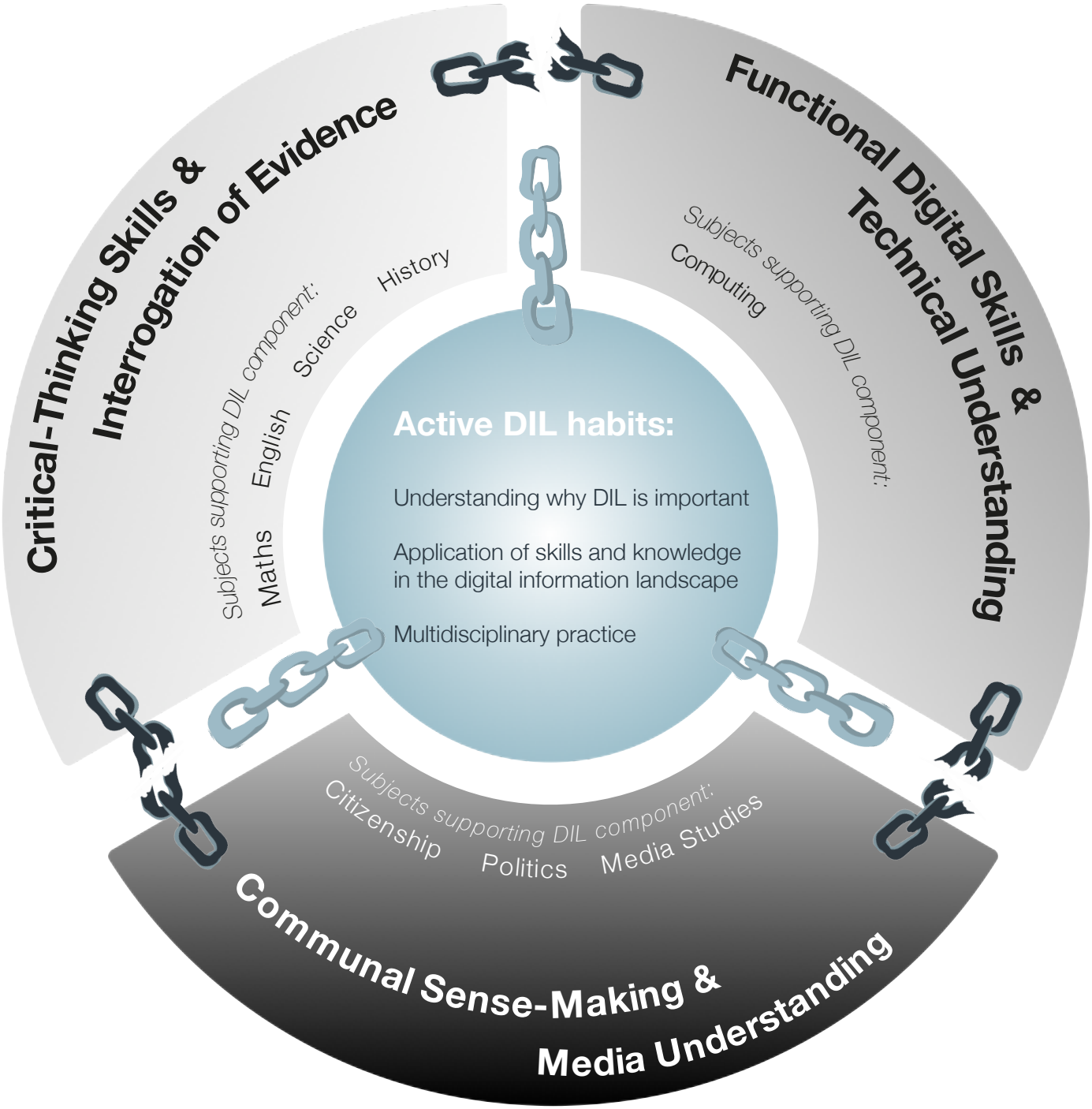
For example, a recent study pointed to surprisingly low levels of DIL among university students and even historians with a PhD, despite their extensive experience as internet users, suggesting that academic excellence and expertise within a subject is not a guarantee of adequate levels of DIL-related skills and knowledge.[100] Teachers should not, therefore, be assumed to possess expertise in DIL by virtue of their high levels of education and subject specialism.

Teacher training is unlikely to have filled such gaps in the DIL levels of teachers, since current teacher-training programmes (such as the PGCE) do not set out specific requirements for or offer training in DIL as a mandatory component of teacher education.

Whilst numerous organisations have created informative DIL-related teaching resources, these cannot act as substitutes for systematic support and training for two reasons. First, the vast majority of such resources and training opportunities focus on the delivery of one-off (or a short series of) sessions focusing on specific issues such as ‘fake news’. As such, they offer little practical pedagogical guidance on how DIL can be embedded into teaching and learning in the long-term and do not help teachers to build the long-lasting multidisciplinary linkages within the curriculum necessary for the effective consolidation of DIL-related skills and knowledge over time. Second, the use of such resources and training opportunities by teachers depends largely on self-referral and therefore depends on teachers’ pre-existing recognition of the importance of DIL in education or awareness of the gaps in one’s own DIL-related competencies. Thus, there is no guarantee that the majority of teachers will actively seek out and translate these opportunities into teaching practice.

In the absence of coherent DIL standards and mandatory training for teachers, the capacity and willingness of individual teachers to foster among their students the full set of skills, knowledge and attitudes encompassed by DIL is likely to be inconsistent at best - or very low at worst. Without the introduction of large-scale systematic DIL-related incentives and support for teachers, unequal provisions of DIL within and between schools emerge as the best case scenario for schools in the UK.

FIGURE 10
The disconnect in DIL skills in the school curriculum



Recommendations

Recognise the strategic importance of DIL

- Scholars, practitioners and regulators must collaborate to create **a standard practical definition of DIL**. Such a framework would need to recognise the complex applied interaction of digital, information, media, political, and critical literacy skills that the modern digital information landscape demands from citizens today.
- Create a **national evidence base** around the agreed definition of DIL

Make DIL a core, cross-subject component in schools

- DIL should not be confined to a single subject and instead should be treated as a continuous element of learning that is reinforced and practised by **all subjects**
- Exam boards should identify **clear opportunities and requirements** for the delivery and practice of DIL within the syllabuses of most core subjects
- DIL should form a key component of **school research projects**, including the EPQ
- The national curriculum must incentivise and set out the opportunities for greater **cross-subject collaboration**

Ensure holistic support for DIL education

- Teachers must be provided with DIL-related CPD opportunities, and DIL must form part of **mandatory training** at the point of 'entry' into teaching for all subjects - for example, as part of PGCE programmes.
- Curricular innovation relating to DIL must **engage all stakeholders**, including teachers, and strive for consensus, dialogue and knowledge exchange

Conclusion

Digital technology is making the world better connected - it has never been cheaper, quicker or easier to locate, create and share information on a mass scale. As a result, the digital age is seeing the overwhelming majority people conduct key aspects of their lives online and for longer periods of time. This increased connectivity and digital presence, however, mean that access to the very core of power in democracies - the citizens - is widening and, if left unchecked, risks exposing democracy to multiple intended and unintended harms.

The multidisciplinary review conducted by the Democracy@Risk Project sought to offer more clarity with respect to two areas of such harm - digital political micro-targeting and online misinformation. As summarised in this report, the extant research and scholarship suggest that these forms of activity generate considerable risks for democracy - consequences that are often, albeit not always, unforeseen by those involved.

Our review found that the biggest sources of risks extend beyond the *persuasive* power of the information that enters citizen consciousness through digital micro-targeting or the eco-system of online misinformation. Elections have been characterised by imperfect information and cognitive biases have clouded the political judgements of citizens long before the digital era, and the digital age has not generated the technology that can pierce through these challenges and mould voter opinions with certainty.

Other processes and principles fundamental to the effective functioning of democracy, however, are more at risk. The first of these is *empowered inclusion*, which requires for all citizens to have equal rights to participate in democratic discourse and decision-making, as well as equal protections of those rights. Digital political micro-targeting, for example, hinders the ability of our democratic political system to bring everyone to the table both unintentionally, by enabling political redlining, and intentionally, by facilitating voter suppression efforts. Online misinformation can push citizens further away by generating disempowering uncertainty that encourages resignation and avoidance of politics.

Second, *collective agenda and will formation* processes are at risk. Greater reliance on digital micro-targeting is likely to result in more fragmented policy promises and emphasis on wedge issues, which divide the electorate and make it difficult to establish political priorities. Online misinformation allows foreign actors entry into public deliberation and facilitates polarization that makes it harder for citizens to empathise with others, and therefore understand and accept the reasons that justify collective judgements.

Third, both phenomena are corrosive to *public trust* in political institutions and office-holders, which can, in the long-term, undermine *democratic legitimacy.*

Narrowing gaps in oversight remains a priority for policy-makers, however, this would not eliminate risk for democracy - digital political micro-targeting generates legal incentives for problematic political behaviour, whilst the effects of misinformation that escapes oversight cannot be easily corrected.

Solutions must therefore reflect the complex nature of these two challenges - the cumulative capabilities of digital micro-targeting and the eco-system of online misinformation can be tamed only through a collaborative, preventative and multi-actor approach in which everyone plays their part in reducing democratic *vulnerabilities*.

Improving DIL requires a national framework for assessment and the maintenance of an evidence base. Instead of being seen as a challenge, DIL should be seen as an opportunity to create a common core theme that strengthens the linkages between skills and subject knowledge, allows for the application of such knowledge in the modern world, and helps to critically empower the voters of tomorrow.

Next Steps: A Note for Researchers

During a special launch event in September 2021, we received valuable feedback and further comments on the themes covered in this report from three leading experts[101] in the fields of political micro-targeting, online misinformation and digital information literacy.

This final section seeks to share with the wider research community the key insights we took away from this event, in the hope that this might help to identify some further priorities for policy-oriented research that build on the work conducted by the Democracy@Risk project thus far.

Recommendations:

1. Assessments of potential harms to democracy should apply democratic theory to identify **a workable, unifying vision of democracy as a political system**. Future research should help policy-makers to relate specific forms of democratic harm (such as 'reduced participation') to this broader concept of democracy and, where appropriate, help weigh these harms against any potential democratic gains generated by the same technologies and processes, in a way that is sensitive to cross-country differences in democratic values and priorities.

2. Future assessments of online harms and how these might be tackled should adopt '**positive**', as well as 'negative', perspectives. Specifically, future work should investigate what constitutes 'best practice' as a way of modelling how digital technology can operate within the parameters of democracy or even *aid* it. One example of this is the potential for digital political micro-targeting to boost inclusion and participation if used to connect in a transparent manner with voters who would otherwise be difficult to reach.

   Thus, in addition to focusing on what tech companies, political campaigns and individuals should *not* do and how problematic actions might be *penalised* or *deterred*, future research should provide answers on what such actors *should* do and how these democracy-promoting practices and technological designs can be *rewarded* by regulators and policy-makers.

3. As governments, regulators and tech companies are beginning to tackle online harms, it is important to **investigate the separate set of threats generated by the very practices and policy responses that emerge initially as a means of resolving the problem**. For example, increased policing and removal of 'harmful' content from online platforms, as well as the broadening of what is meant by the term 'harmful' itself, elicit considerable problems for free speech and equity. Research should focus on assessing the potentially threatening consequences of these solutions, which pose further questions regarding who should be making decisions about what is sufficiently harmful, who should have the right to remove content, and who watches those arbiters.

4. Future research should **explore the variation in democratic and online vulnerabilities across different country and cultural contexts**. Whilst some countries and communities might be more vulnerable than others overall, it is important to note that different countries and communities can be resilient and vulnerable to different aspects of the same problem.

Appendix A: Method

This report highlights the key policy-oriented conclusions of a narrative literature review conducted by our team. Our aim was to review and synthesise the extant research to generate a more holistic understanding of three areas of potential risk for democracy - digital political micro-targeting, online misinformation and digital information literacy.

For each of the three concepts under investigation, the review was guided by the following questions:

1. *Building definition*: what does this practice/phenomenon entail and how does it function?

2. *Understanding risk*: does this practice/phenomenon pose a challenge for democracy and, if so, what is the nature of this challenge?

3. *Towards solutions:* what do the answers to the above questions mean for how these challenges should be tackled?

The initial literature search strategy comprised multiple search methods, including electronic database searches (JSTOR, Sage Journals, Science Direct, Scopus, Taylor & Francis Online, University of Manchester Library, Web of Science, Wiley Online Library), hand searches of peer-reviewed journals (Journal of Democracy; Internet Policy Review) and ancestry searches. From this search, we selected English-language peer-reviewed studies (theoretical or empirical research deploying quantitative, qualitative or mixed-method research designs) and 'grey' literature, published between 2005 and 2021.

The initial findings of the review were presented at three academic workshops, each dedicated to one of the three challenge areas under review. The discussions at these events provided valuable feedback and generated additional literature for the review on the basis of the recommendations made by workshop participants.

Finally, at a special launch event in September 2021, we received further comments on the report from three experts, each working on one of the three challenge areas under review, which were used to refine the report contents further and informed the final section of the report discussing the possible next steps for research.

References

1. Ofcom. (2020). Online Nation 2020 Report. 24 June 2020. [Online]. Available from https://www.ofcom.org.uk/__data/assets/pdf_file/0027/196407/online-nation-2020-report.pdf [Accessed June 2021]

2. Foa, R.S., Klassen, A., Slade, M., Rand, A. and R. Collins. (2020). *The Global Satisfaction with Democracy Report 202*0. January 2020. Cambridge, United Kingdom: Centre for the Future of Democracy. Available from https://www.bennettinstitute.cam.ac.uk/media/uploads/files/DemocracyReport2020.pdf [Accessed June 2021]

3. Baldwin-Philippi, K. (2017). The myths of data driven campaigning, *Political Communication*, *34*(4). 627–633.
Baldwin-Philippi, J. (2019). Data campaigning: Between empirics and assumptions, *Internet Policy Review*, *8*(4). 1-18.

4. The Electoral Commission. (2019). Digital campaigning - increasing transparency for voters. 13 August 2019. [Online]. Available from https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/changing-electoral-law/transparent-digital-campaigning/report-digital-campaigning-increasing-transparency-voters [Accessed June 2021]

5. Wagner, K., Bergen, M., & Frier, S. (2020). 'Big tech draws record revenue, harsh criticism with election ads', *Bloomberg (Online)*. 2 November. Available from https://www.bloomberg.com/news/articles/2020-11-02/big-tech-draws-record-revenue-harsh-criticism-with-election-ads [Accessed June 2021]

6. Frier, S. (2018). 'Trump's campaign said it was better at Facebook. Facebook agrees', *Bloomberg (Online).* 3 April. Available from https://www.bloomberg.com/news/articles/2018-04-03/trump-s-campaign-said-it-was-better-at-facebook-facebook-agrees [Accessed June 2021]

7. Google. (2021). Transparency report. 21 March 2019 - 1 June 2021. [Online]. Available from https://transparencyreport.google.com/political-ads/region/EU [Accessed June 2021]

8. Rentfrow, P. J., Jost, J. T., Gosling, S. D., & Potter, J. (2009). 'Statewide differences in personality predict voting patterns in 1996-2004 U.S. Presidential elections', in Jost, J.T., Kay, A.C. & Thorisdottir, H. (eds.), *Series in political psychology. Social and psychological bases of ideology and system justification*. Oxford University Press. 314–347
Mondak, J. (2010) *Personality and the foundations of political behavior* (Cambridge studies in public opinion and political psychology). Cambridge: Cambridge University Press.
Johnston, C., Lavine, H., & Federico, C. (2017). *Open versus closed: personality, identity, and the politics of redistribution*. Cambridge: Cambridge University Press.

9. Dubois, D., Rucker, D., & Galinsky, A. (2016). Dynamics of communicator and audience power: The persuasiveness of competence versus warmth. *Journal of Consumer Research*, *43*(1). 68-85.
Matz, S., Kosinski, M., Nave, G., & Stillwell, D. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences of the United States*, *114*(48). 12714-12719.
Moon, Y. (2002). Personalization and personality: Some effects of customizing message style based on consumer personality. J*ournal of Consumer Psychology*, *12*(4). 313-325.
Hirsh, J., Kang, S., & Bodenhausen, G. (2012). Personalized persuasion: tailoring persuasive appeals to recipients' personality traits. *Psychological Science*, *23*(6). 578-581.
Wheeler, S., Petty, R., & Bizer, G. (2005). Self-schema matching and attitude change: Situational and dispositional determinants of message elaboration. *Journal of Consumer Research*, *31*(4). 787-797.

10. Lock, A., & Harris, P. (1996). Political marketing - vive la différence! *European Journal of Marketing*, *30*(10/11). 14-24.
Peng, N., & Hackley, C. (2009). Are voters consumers? A qualitative exploration of the voter-consumer analogy in political marketing, *Qualitative Market Research, 12 (*2). 171–86.
Gelb, B.D., & Bush, D. (2011). Advertising and policy insights for the voter versus customer trade-off. *Journal of Public Policy & Marketing, 30*(1). 96–99.

11. Coppock, A., Hill, S.J., & Vavreck, L. (2020). The small effects of political advertising are small regardless of context, message, sender, or receiver: Evidence from 59 real-time randomized experiments. *Science Advances,* 6(36). eabc4046.

12. Boerman, S., & Kruikemeier, S. (2016). Consumer responses to promoted tweets sent by brands and political parties. *Computers in Human Behavior,* 65. 285-294.

13. Dommett, K. (2019). Data-driven political campaigns in practice: understanding and regulating diverse data-driven campaigns. *Internet Policy Review*, 8(4).

14. Schumann, S. & Klein, O. (2015). Substitute or stepping stone? Assessing the impact of low-threshold online collective actions on offline participation. *European Journal of Social Psychology*. 45(3). 308-322.

15. Bennett, C. (2016). Voter databases, micro-targeting, and data protection law: can political parties campaign in Europe as they do in North America?, *International Data Privacy Law*, *6*(4). 261–275.

16. Ali, M., Sapiezynski, P., Korolova, A., Mislove, A. & Rieke, A. (2021). Ad delivery algorithms: the hidden arbiters of political messaging. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM '21)*. Association for Computing Machinery. New York, USA. 13–21.

17. Hersch, E., & Schaffer, B. (2013). Targeted campaign appeals and the value of ambiguity, *The Journal of Politics, 75*(2). 520-534.
Flores, A., & Coppock, A. (2018). Do bilinguals respond more favorably to candidate advertisements in English or in Spanish?, *Political Communication, 35*(4), 612-633.

18. Hersch, E. (2015). *Hacking the electorate: how campaigns perceive voters.* New York: Cambridge University Press.

19. Nielsen, R. (2012). *Ground Wars*. Princeton: Princeton University Press
Baldwin-Philippi, K. (2017). The myths of data driven campaigning, *Political Communication*, *34*(4). 627–633.
Dommett, K. (2019). Data-driven political campaigns in practice: Understanding and regulating diverse data-driven campaigns. *Internet Policy Review*, 8(4).

20. APPG on Electoral Campaigning Transparency. (2020). Defending our democracy in the digital age. [Online]. Available from https://fairvote.uk/defending-our-democracy-in-the-digital-age-new-report-launched/ [Accessed April 2020].
The Electoral Commission. (2021). Reforming electoral law. *Official Website of The Electoral Commission.* [Online]. Available from https://www.electoralcommission.org.uk/sites/default/files/pdf_file/Reforming-electoral-law-PACAC-booklet.pdf [Accessed June 2021]

21. Bennett, C. (2016). Voter databases, micro-targeting, and data protection law: can political parties campaign in Europe as they do in North America?, *International Data Privacy Law*, *6*(4). 261–275.

22. Privacy International. (2018). UK Data Protection Act 2018–339 pages still falls short on human rights protection, *Privacy International (Online)*. 13 June. Available from https://privacyinternational.org/news-analysis/2074/uk-data-protection-act-2018-339-pages-still-falls-short-human-rights-protection [Accessed June 2020]

23. Hern, A. & McIntyre, N. (2019). Google admits major underreporting of election ad spend, *The Guardian (Online).* 19 November. Available from https://www.theguardian.com/technology/2019/nov/19/google-admits-major-underreporting-of-election-ad-spend [Accessed June 2021]

24. Scott, M. (2019). Political ads on Facebook disappear ahead of UK election, *Politico (Online).* 10 December. Available from https://www.politico.com/news/2019/12/10/political-ads-on-facebook-disappear-ahead-of-uk-election-081376 [Accessed June 2021]
Manthorpe, R. (2019). Researchers fear 'catastrophe' as political ads 'disappear' from Facebook library, *Sky News*, 11 December. Available from https://news.sky.com/story/researchers-fear-catastrophe-as-political-ads-disappear-from-facebook-library-11882988 [Accessed May 2020]

25. Ali, M., Sapiezynski, P., Korolova, A., Mislove, A. & Rieke, A. (2021). Ad delivery algorithms: the hidden arbiters of political messaging. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM '21)*. Association for Computing Machinery. New York, USA. 13–21.

26. Libert, T. (2018). An automated approach to auditing disclosure of third-party data collection in website privacy policies. *The 2018 Web Conference.* April 23-27, 2018. Lyon, France.

27. Costa, E. & Halpern, D. (2019). The behavioural science of online harm and manipulation, and what to do about it, *The Behavioural Insights Team,* March 2019. [Online] Available from https://www.bi.team/wp-content/uploads/2019/04/BIT_The-behavioural-science-of-online-harm-and-manipulation-and-what-to-do-about-it_Single.pdf [Accessed April 2020]

28. Libert, T. (2018). An automated approach to auditing disclosure of third-party data collection in website privacy policies. *The 2018 Web Conference.* April 23-27, 2018. Lyon, France.
as cited in Costa, E. & Halpern, D. (2019). The behavioural science of online harm and manipulation, and what to do about it, *The Behavioural Insights Team,* March 2019. [Online] Available from https://www.bi.team/wp-content/uploads/2019/04/BIT_The-behavioural-science-of-online-harm-and-manipulation-and-what-to-do-about-it_Single.pdf [Accessed April 2020]

29. Open Rights Group. (2020). Public are kept in the dark over data driven political campaigning, poll finds. *Open Rights Group (Online Press Release).* 10 January. Available from https://www.openrightsgroup.org/press-releases/public-are-kept-in-the-dark-over-data-driven-political-campaigning-poll-finds/ [Accessed June 2021]

30. Mozilla. (2019). Data Collection Log — EU Ad Transparency Report, *Mozilla (Online).* Available from https://adtransparency.mozilla.org/eu/log/ [Accessed May 2020]
Mozilla. (2019). Facebook and Google: This is What an Effective Ad Archive API Looks Like, *Mozilla (Online).* Available from https://blog.mozilla.org/en/mozilla/facebook-and-google-this-is-what-an-effective-ad-archive-api-looks-like/ [Accessed June 2021]
Who Targets Me. (2020).How to take a "gold standard" approach to political advertising transparency and policy, *Who Targets Me (Online).* Available from https://whotargets.me/en/how-to-take-a-gold-standard-approach-to-political-advertising-transparency-and-policy/ [Accessed June 2021]

31. Facebook. (2019). Boost liquidity and work smarter with machine learning, *Facebook,* 27 March. [Online] Available from https://www.facebook.com/business/news/insights/boost-liquidity-and-work-smarter-with-machine-learning [Accessed June 2020].

32. Costa, E. & Halpern, D. (2019). The behavioural science of online harm and manipulation, and what to do about it, *The Behavioural Insights Team,* March 2019. [Online] Available from https://www.bi.team/wp-content/uploads/2019/04/BIT_The-behavioural-science-of-online-harm-and-manipulation-and-what-to-do-about-it_Single.pdf [Accessed April 2020]

33. ICO. (2018). Democracy disrupted: Personal information and political influence. July 2018. [Online]. Available from https://ico.org.uk/media/action-weve-taken/2259369/democracy-disrupted-110718.pdf [Accessed June 2020]

34. Howard, P. (2006). *New media and the managed citizen.* New York: Cambridge University Press.

35. Endres, K., & Kelly, K. (2018). Does microtargeting matter? Campaign contact strategies and young voters, *Journal of Elections, Public Opinion and Parties*, *28*(1). 1-18.

36. Nickerson, D. W. & Rogers, T. (2014). Political campaigns and Big Data, *Journal of Economic Perspectives. 28* (2). 51–74.

37. Green, J. & Issenberg, S. (2016). Inside the Trump bunker, with days to go, *Bloomberg (Online).* 27th October. Available from https://www.bloomberg.com/news/articles/2016-10-27/inside-the-trump-bunker-with-12-days-to-go [Accessed April 2020].

38. Green, D., Mcgrath, M., & Aronow, P. (2013). Field Experiments and the Study of Voter Turnout. Journal of Elections, *Public Opinion and Parties*, *23*(1). 27-48.
Townsley, J. (2018). Is it worth door-knocking? Evidence from a United Kingdom-based Get Out The Vote (GOTV) field experiment on the effect of party leaflets and canvass visits on voter turnout, *Political Science Research and Methods.* 1-15.

39. Bizzotto, J. & Solow, B. (2019) Electoral Competition with Strategic Disclosure, *Games*, *10*(3).

40. Hillygus, D.S. & Shields, T.G. (2008). *The persuadable voter: Wedge issues in presidential campaigns*. Princeton University Press.

41. Glaeser, E.L., Ponzetto, G.A.M, & Shapiro, J.M. (2005). Strategic extremism: Why Republicans and Democrats divide on religious values, *The Quarterly Journal of Economics. 120*(4). 1283-1330.

42. Examples of more detailed recommendations for ad library standards can be found in:
Who Targets Me. (2020).How to take a “gold standard” approach to political advertising transparency and policy, *Who Targets Me (Online).* Available from https://whotargets.me/en/how-to-take-a-gold-standard-approach-to-political-advertising-transparency-and-policy/
Mozilla. (2019). Facebook and Google: This is What an Effective Ad Archive API Looks Like, *Mozilla (Online).* Available from https://blog.mozilla.org/en/mozilla/facebook-and-google-this-is-what-an-effective-ad-archive-api-looks-like/ [Accessed June 2021]

43. More detailed recommendations for auditing tools can be found in:
Who Targets Me. (2020).How to take a “gold standard” approach to political advertising transparency and policy, *Who Targets Me (Online).* Available from https://whotargets.me/en/how-to-take-a-gold-standard-approach-to-political-advertising-transparency-and-policy/ [Accessed June 2021]

44. More detailed recommendations on increasing financial sanctions can be found in:
APPG on Electoral Campaigning Transparency. (2020). Defending our democracy in the digital age. [Online]. Available from https://fairvote.uk/defending-our-democracy-in-the-digital-age-new-report-launched/ [Accessed April 2020].
The Electoral Commission. (2021). Reforming electoral law. *Official Website of The Electoral Commission.* [Online]. Available from https://www.electoralcommission.org.uk/sites/default/files/pdf_file/Reforming-electoral-law-PACAC-booklet.pdf [Accessed June 2021]

45. Innes, M. (2020). Techniques of disinformation: Constructing and communicating “soft facts” after terrorism, *The British Journal of Sociology, 71*(2). 284-299.

46. Nielsen, R.K., Fletcher, R., Newman, N., Brennen, J.S., & Howard, P. (2020). Navigating the ‘Infodemic’: How People in Six Countries Access and Rate News and Information about Coronavirus. Reuters Institute for the Study of Journalism, April 2020. Available from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-04/Navigating%20the%20Coronavirus%20Infodemic%20FINAL.pdf [Accessed May 2020]

47. Innes, M. (2020). Techniques of disinformation: Constructing and communicating “soft facts” after terrorism, *The British Journal of Sociology, 71*(2). 284-299.

48. Howard et al. (2018). The IRA and Political Polarization in the United States, *Oxford Internet Institute*, August 22, 2018. Available from: https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/12/The-IRA-Social-Media-and-Political-Polarization.pdf [Accessed April 2020].
DiResta, R., et al. (2018). The Tactics & Tropes of the Internet Research Agency, *New Knowledge*, December 17, 2018. Available from https://cdn2.hubspot.net/hubfs/4326998/ira-report-rebrand_FinalJ14.pdf [Accessed 10 April 2020]

49. Nimmo, B., Eib, S., Ronzaud, L., Ferreira, R., Lederer, T., & Smith, M. (2020). Iran’s Broadcaster: Inauthentic Behavior*, Graphika Report*, 5th May 2020. Available from https://www.graphika.com/reports/irans-broadcaster-inauthentic-behavior/ [Accessed May 2020]

50. Nimmo, B., Francois, C., Eib, S. & Ronzaud, L. (2020). Return of the (Spamouflage) Dragon, *Graphic Report*, 24th April 2020. Available from https://www.graphika.com/reports/return-of-the-spamouflage-dragon-1/ [Accessed May 2020]
Twitter Safety. (2019). Information operations directed at Hong Kong, *Twitter Safety*, 19th August 2019. Twitter. Available from https://blog.twitter.com/en_us/topics/company/2019/information_operations_directed_at_Hong_Kong.html [Accessed April 2020]

51. Keller, F., Schoch, D., Stier, S. & Yang, J. (2020). Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign, *Political Communication*, 37(2). 256-280.

52. Oates, S. (2018). Projecting power: understanding Russian strategic narrative, *Russian Analytical Digest,* 229. Available from https://ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/RAD229.pdf [Accessed April 2020]

53. Innes, M. (2020). Techniques of disinformation: Constructing and communicating “soft facts” after terrorism, *The British Journal of Sociology, 71*(2). 284-299.

54. Subramanian, S. (2017). Inside the Macedonian fake-news complex, *Wired*, 15 February. Available from https://www.wired.com/2017/02/veles-macedonia-fake-news/ [Accessed April 2020]

55. Townsend, T. (2016). The Bizarre Truth Behind the Biggest Pro-Trump Facebook Hoaxes, *Inc*, 21 November. Available from https://www.inc.com/tess-townsend/ending-fed-trump-facebook.html [Accessed April 2020]

56. Dewey, C. (2014).This is not an interview with Banksy, *The Washington Post (Online)*, 22 October. Available from https://www.washingtonpost.com/news/the-intersect/wp/2014/10/21/this-is-not-an-interview-with-banksy/ [Accessed April 2020]
Sydell, L. (2016). We Tracked Down A Fake-News Creator In The Suburbs. Here's What We Learned, *National Public Radio (Online)*, 23 November. Available from https://www.npr.org/sections/alltechconsidered/2016/11/23/503146770/npr-finds-the-head-of-a-covert-fake-news-operation-in-the-suburbs?t=1589356235490&t=1623218118291 [Accessed April 2020]

57. Rid, T. (2019). Who's Really to Blame for the 'Ukraine Did It' Conspiracy Theory?, *The Atlantic (Online)*, 5th December. Available from https://www.theatlantic.com/ideas/archive/2019/12/who-created-ukraine-did-it-conspiracy-theory/602992/ [Accessed April 2020]

58. Centre for Countering Digital Hate. (2021). The Disinformation Dozen: Why platforms must act on twelve leading online anti-vaxxers. *Centre for Countering Digital Hate (Online).* Available from https://252f2edd-1c8b-49f5-9bb2-cb57bb47e4ba.filesusr.com/ugd/f4d9b9_b7cedc0553604720b7137f8663366ee5.pdf [Accessed June 2021]

59. Nimmo, B., François, C., Shawn Eib, C., Ronzaud, L. & Carter, J. (2020). GRU and the minions: Further exposures of Russian military assets across platforms, 2013-2020. Graphika. [Online]. Available from https://graphika.com/reports/gru-and-the-minions/ [Accessed August 2021]
Nimmo, B., François, C., Shawn Eib, C., Ronzaud, L., Ferreira, R., Hernon, C. & Kostelancik, T. (2020). Exposing Secondary Infektion. Graphika. [Online]. Available from https://graphika.com/reports/exposing-secondary-infektion/ [Accessed August 2021]
Barash, V. (2020). Anatomy of a Disinformation Campaign, presented at *Combatting misinformation online: The role and relevance of the social sciences, Aspect Annual Event*. 24 September. University of Manchester. [Online].

60. Keller, F., Schoch, D., Stier, S. & Yang, J. (2020). Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign, *Political Communication*, 37(2). 256-280.

61. Annual figure calculated using the quarterly figures reported in Facebook's Transparency Center. Available from https://transparency.fb.com/data/community-standards-enforcement/fake-accounts/facebook#CONTENT_ACTIONED [Accessed June 2021]

62. Twitter Safety. (2020). Disclosing networks of state-linked information operations we've removed, *Twitter Safety*, 11th June. Twitter. Available from https://blog.twitter.com/en_us/topics/company/2020/information-operations-june-2020.html [Accessed June 2020]

63. Institute For The Future Digital Intelligence Lab. (2019) False information in the current news environment. [Online Brief]. Available from https://www.iftf.org/fileadmin/user_upload/images/DigIntel/1_False_information_in_current_news_FINAL_031119.pdf [Accessed June 2021]

64. Institute For The Future Digital Intelligence Lab. (2019) Mitigating the negative impact of false information. [Online Brief]. Available from https://www.iftf.org/fileadmin/user_upload/images/DigIntel/3_Mitigating_Negative_impact_of_False_information_FINAL_031119.pdf [Accessed June 2021]

65. Nielsen, R.K., Fletcher, R., Newman, N., Brennen, J.S., & Howard, P. (2020). Navigating the 'Infodemic': How People in Six Countries Access and Rate News and Information about Coronavirus. Reuters Institute for the Study of Journalism, April 2020. Available from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-04/Navigating%20the%20Coronavirus%20Infodemic%20FINAL.pdf [Accessed May 2020]

66. Chadwick, A., Vaccari, C. & O'Loughlin, B. (2018). Do tabloids poison the well of social media? Explaining democratically dysfunctional news sharing, *New Media & Society, 20*(11). 4256-4274.
Govolchenko, Y., Hartmann, M., & Adler-Nissen, R. (2018). State, media and civil society in the information warfare over Ukraine: citizen curators of digital disinformation, *International Affairs, 94*(5). 975–994.

67. Singer, J. (2014). User-generated visibility: Secondary gatekeeping in a shared media space, *New Media & Society, 16*(1). 55-73.

68. Innes, M. (2020). Techniques of disinformation: Constructing and communicating "soft facts" after terrorism, *The British Journal of Sociology, 71*(2). 284-299.

69. Grinberg, N., Joseph, K., Friedland, L, Swire-Thomp- son, B., & Lazar, D. (2019). Fake news on Twitter during the

71. Huang, H. (2017). A War of (Mis)Information: The political effects of rumors and rumor rebuttals in an authoritarian country, *British Journal of Political Science. 47*(2). 283–31.

72. McKay, S., & Tenove, C. (2020). Disinformation as a threat to deliberative democracy, *Political Research Quarterly.* 1. 1-15.

73. Mitchell, A., Gottfried, J., Stocking, G., Walker, M. & Fedeli, S. (2019). Many Americans Say Made-Up News Is a Critical Problem That Needs To Be Fixed, *Pew Research Center (Online).* 5 June. Available from https://www.journalism.org/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/ [Accessed June 2021]

74. Arif, A., Stewart, L. & Starbird, K. (2018). Acting the Part: Examining Information Operations Within #BlackLivesMatter Discourse, *Proceedings of the ACM on Human-Computer Interaction.* 2, CSCW, Article 20.

75. Ognyanova, K., Lazer, D., Robertson, R. & Wilson, C. (2020). Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power, *The Harvard Kennedy School (HKS) Misinformation Review.* 1(4).

76. Nyhan, B., Reifler, J., & Ubel, P. (2013). The hazards of correcting myths about health care reform, *Medical Care.* 51(2). 127-132.
Nyhan, B., Porter, E., Reifler, J., & Wood, T. J. (2019). Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability, *Political Behavior*. 1-22.
Weeks, B. (2015). Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation, *Journal of Communication*, *65*(4). 699-719.
Porter, E., Wood, T., & Kirby, D. (2018). Sex trafficking, Russian infiltration, birth certificates, and pedophilia: A survey experiment correcting fake news, *Journal of Experimental Political Science*, *5*(2). 159-164.
Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence, *Political Behavior*, *41*(1). 135-163.

77. Jarman, J. (2016). Influence of political affiliation and criticism on the effectiveness of political fact-checking, Communication Research Reports, 33(1). 9-15.

78. Fridkin, K., Courey, J., Hernandez, S., & Spears, J. (2016). Gender differences in reactions to fact checking of negative commercials, *Politics & Gender, 12*(2). 369-390.

79. Fridkin, K., Kenney, P., & Wintersieck, A. (2015). Liar, liar, pants on fire: How fact-checking influences citizens' reactions to negative advertising, *Political Communication*, *32*(1). 127-151.

80. Nyhan, B., Reifler, J., & Ubel, P. (2013). The hazards of correcting myths about health care reform, *Medical Care.* 51(2). 127-132.

81. Swire-Thompson, B., Ecker, U., Lewandowsky, S., & Berinsky, A. (2020). They might be a liar but they're my liar: Source evaluation and the prevalence of misinformation, *Political Psychology*, *41*(1). 21-34.
Nyhan, B., & Zeitzoff, T. (2017). Fighting the past: Perceptions of control, historical misperceptions, and corrective information in the Israeli-Palestinian conflict, *Political Psychology, 39*(3). 611–631.
Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. (2017). Processing political misinformation: Comprehending the Trump phenomenon, *Royal Society Open Science, 4*(3). 16080.
Nyhan, B., Porter, E., Reifler, J., & Wood, T. J. (2019). Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability, *Political Behavior*. 1-22.

82. Anderson, C. A., Lepper, M., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information, J*ournal of Personality and Social Psychology*, 396. 1037–1049.
Anderson, C. A., New B. L., & Speer, J. R. (1985). Argument availability as a mediator of social theory perseverance, *Social Cognition*, *3*(3). 1235–1249.
Ecker, U. K., Lewandowsky, S., Swire, B., & Chang, D. (2011). Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction, *Psychonomic Bulletin and Review*, *18*(3). 570–578.
Ecker, U. K., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation, *Memory and Cognition*, *38*(8). 1087– 1100.

83. Thorson, E. (2016). Belief Echoes: The Persistent Effects of Corrected Misinformation. Political Communication, 33(3), 460-480.

84. McGuire, W. J. (1961). The effectiveness of supportive and refutational defenses in immunizing and restoring beliefs against persuasion, *Sociometry*, *24.* 184–197.
McGuire, W. J. & Papageorgis, D. (1961). The relative efficacy of various types of prior belief-defense in producing immunity against persuasion, J*ournal of Abnormal Social Psychology, 62.* 327–337.

85. Compton, J. & Pfau, M. (2005). Inoculation theory of resistance to influence at maturity: Recent progress in theory development and application and suggestions for future research, in Kalbfleisch, P. (ed.) *Communication yearbook* 29. 97–145. Newbury Park, CA: Sage.
Banas, J., & Miller, G. (2013). Inducing Resistance to Conspiracy Theory Propaganda: Testing Inoculation and Metainoculation Strategies, *Human Communication Research*, *39*(2). 184-207.

86. Banas, J. & Rains, S. (2010). A Meta-Analysis of Research on Inoculation Theory, *Communication Monographs*, 77(3). 281-311.

87. Cook, J., Lewandowsky, S., & Ecker, U. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence, *PLoS ONE*. 12(5). E0175799.
Jolley, D. & Douglas, K.M. (2017). Prevention is better than cure: addressing antivaccine conspiracy theories. *Journal of Applied Social Psychology*.

88. Roozenbeek, J., & van der Linden, S. (2018). The fake news game: actively inoculating against the risk of misinformation, *Journal of Risk Research*, *22*(5). 570–580.
Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation, *Nature Palgrave Communications*, *5*(65).
Cook, J., Lewandowsky, S., & Ecker, U. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence, *PLoS ONE*. 12(5). E0175799.

89. Banas, J., & Miller, G. (2013). Inducing resistance to conspiracy theory propaganda: Testing inoculation and metainoculation strategies, *Human Communication Research*, *39*(2). 184-207.

90. Digital, Culture, Media and Sports Committee. (2019). *Disinformation and 'fake news': Final report.* (HC 1791, 2017-19). [Online]. London: House of Commons. Available from https:// publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf [Accessed April 2020]

91. House of Lords Select Committee on Democracy and Digital Technologies. (2020). *Digital Technology and the Resurrection of Trust*. (HL Paper 77, 2019-2021). London: House of Lords. Available from: https://committees.parliament.uk/publications/1634/documents/17731/default/ [Accessed February 2021]

92. Kahne, J., & Bowyer, B.T. (2017). Educating for democracy in a partisan age, *American Educational Research Journal*, *54*. 3-34.

93. Nsangi, A., Semakula, D., Oxman, A. D., Austvoll-Dahlgren, A., Oxman, M., Rosenbaum, S., . . . Sewankambo, N. K. (2017). Effects of the Informed Health Choices primary school intervention on the ability of children in Uganda to assess the reliability of claims about treatment effects: A cluster- randomised controlled trial, *The Lance*t, *390*(10092). 374-388.

94. Lessenski, M. (2021). Media Literacy Index 2021. Double Trouble: Resilience to Fake News at the Time of Covid-19 Infodemic, *Open Society Institute Sofia (Online).* Available from https://osis.bg/wp-content/uploads/2021/03/MediaLiteracyIndex2021_ENG.pdf [Accessed June 2021]

95. Halinen, I. (2018). The new educational curriculum in Finland, in Matthes,M, Pulkkinen, L., Clouder, C., & Hey, B. (ed.) *Improving the Quality of Childhood in Europe vol.7.* Brussels, Belgium: Alliance for Childhood European Network Foundation. 75-89. Available from https://www.allianceforchildhood.eu/files/Improving_the_quality_of_Childhood_Vol_7/QOC%20V7%20CH06%20DEF%20WEB.pdf [Accessed June 2021]

96. National Literacy Trust. (2018). *Fake news and critical literacy: The final report of the Commission on Fake News and the Teaching of Critical Literacy in Schools.* [Online] London: National Literacy Trust. Available from https://cdn.literacytrust.org.uk/media/documents/Fake_news_and_critical_literacy_-_final_report.pdf [Accessed April 2020]

97. Digital, Culture, Media and Sports Committee. (2018). *Disinformation and 'fake news': Interim report: Government Response to the Committee's Fifth Report of Session 2017–19.* (HC 1640, 2017-2019). [Online]. London: House of Commons. Available from https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1630/1630.pdf [Accessed February 2021]

House of Lords Select Committee on Democracy and Digital Technologies. (2020). *Government Response to the House of Lords Democracy and Digital Technologies Committee Report on Digital Technology and the Resurrection of Trust.* (CP 285). [Online]. London: House of Lords. Available from https://committees.parliament.uk/publications/2308/documents/22803/default/ [Accessed February 2021].

98. Polizzi, G. (2020). Digital literacy and the national curriculum for England: Learning from how the experts engage with and evaluate online content, *Computers & Education*, 152. 103859.

99. Larke, L. (2019). Agentic neglect: Teachers as gatekeepers of England's national computing curriculum. *British Journal of Educational Technology*, *50*(3), 1137–1150.

100. Wineburg, S. & McGrew, S. (2017). Lateral reading: Reading less and learning more when evaluating digital information. Working Paper No. 2017.A1. Palo Alto: Stanford History Education Group, Stanford University. Available from: https://purl.stanford.edu/yk133ht8603 [Accessed June 2020]

101. Dr Kate Dommett (University of Sheffield), Professor Martin Innes (Cardiff University Crime and Security Research Institute) and Dr Kari Kivinen (EUIPO Observatory). 20 September 2021.