DEMOCRACY & DISINFORMATION: A TURN IN THE DEBATE

KVAB Thinkers' Report

Anja Bechmann

Ben O'Loughlin

Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten
Paleis der Academiën – Hertogsstraat 1 – 1000 Brussel – België
info@kvab.be – www.kvab.be

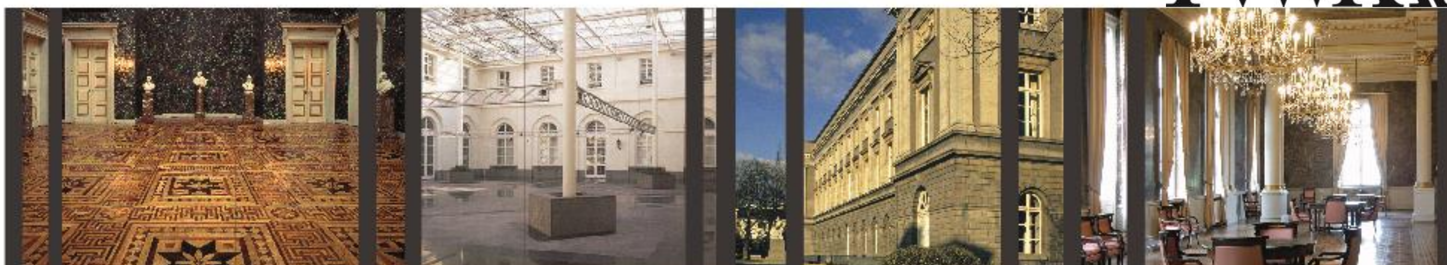

TABLE OF CONTENTS

DEMOCRACY & DISINFORMATION: A TURN IN THE DEBATE

0. Introduction 5

1. Disinformation & Democracy: A multi-layered debate 8

2. The Flemish context: Cracks in the system? 18

3. A turn in the debate: Existing recommendations and our proposals 21

Evaluation of existing recommendations 21

Concluding recommendations and future directions 27

4. Conclusion 31

5. References 33

0. Introduction

The purpose of this thinkers' report is to establish what existing research and reports about disinformation and democracy tell us and to evaluate current policy recommendations in order to establish what might be feasible evidence-based proposals. The report is not the result of a negotiated outcome between different stakeholders. Instead, it aims to make clear what research has been done and not done, on this topic in order to re-direct the parameters of policy options towards those based on substantive research rather than conjecture about how disinformation is presumed to work.

Disinformation and democracy, and related debates about fake news and post-truth knowledge, are highly contested and political topics that involve multiple stakeholders. We acknowledge and believe that stakeholders need to come together to understand this complex phenomenon. At the same time, however, we want to strengthen historical, theoretical and evidence-based discussions on the topic. This report has three objectives. First, it seeks to establish what we know and what we do not know about disinformation. Second, it identifies what from a research point of view specifically it is important to pursue in future studies. Third, it identifies what policies and regulation follow from that knowledge that could address the current dilemmas faced by democracies. This report addresses this at the national and subnational level, with a particular focus on Belgium and the Flemish region, and at a European and international level.

Due to the encompassing character of the topic our ambition is not to lay out all research done on this topic. Instead, and supplementing the KVAB position paper on fake news and disinformation (Billiet, Opgenhaffen, Pattyn, & Van Aelst, 2019), we will focus in particular on social media in the context of disinformation because disinformation is frequently forwarded via social media. However, we – and the research we review -- also acknowledge that disinformation exists across all media and across many contexts. The overarching question then becomes: What are the potential consequences (if any) of this sharing behavior for the functioning of democracy? In order to answer this very big question we will outline key dilemmas in the context of this question by focusing on challenges with conceptualizing the phenomenon and detecting and mitigating the logics of spread and not least evaluating the danger for democracy. Our focal point will be the European context and specifically minority language areas such as the Flemish region of Belgium, but this will be done as perspectives to and in the context of existing studies that primarily are to be found in the U.S. and British context.

Flanders is an important case. Across all 28 member states, citizens of Belgium per se are least likely to say false news is a problem in their country (European Commission, 2018). When we interviewed experts in early 2019, they felt that polarisation was increasing but that voters generally

felt a high degree of trust in mainstream media and democratic institutions. However, May 2019 federal elections saw a far-right party surge from relative obscurity to take second place at the polls after spending significant sums on campaigning (*The Brussels Times*, 2019). Is this small corner of European prosperity, stability and security succumbing to wider forces across all democracies, and what can we learn from this experience?

Methodology

The topic of "Disinformation & Democracy" as one of the annual topics of KVAB Thinker's Programme in 2019 allowed for us to build on existing associated topics in our research on strategic narratives (Miskimmon et al., 2013, 2017) and filter bubbles (Bechmann & Nielbo, 2018). In particular, we extended our research to the analysis and contextualization of current data exchange solutions and policies (Moeller & Bechmann, 2019) and the impact of social media on democratic participation (Chadwick et al., 2015, 2018) that are of particular interest and a prerequisite when researching disinformation. The Thinker's programme allowed for us to carry out interviews with key stakeholders in Brussels and participate in important stakeholder meetings to better understand incentives, needs, and hopes. We also held meetings with Flemish and Belgian research colleagues in order to better understand the Flemish political climate and the presence of disinformation there. Alongside our own research backgrounds, data from these meetings and interview form the basis for the discussions in this report. The meetings and interviews were held during Spring and Summer 2019 and consisted among others of two fact checking weeks in Brussels in February and April. We met with Flemish journalists from the largest Flemish media, and researchers from IMEC, VUB, KU Leuven, the University of Gent, and the University of Antwerp. In April we participated and presented at the KU Leuven expert seminar 'Media, Post-Truth and the Future of Democracy' with the fellows of Metaforum – Nicole Curato, Vibodh Parthasarathi, and Michael Delli Carpini. During the Spring and Summer, we furthermore supplemented with interviews and meetings with platform representatives, ministries, NGOs, foundations, ambassadors, and commission representatives. The concluding symposium was held in the Flemish Parliament, October 11, 2019.

This report has three parts. In the first, we outline the key components of the debate about disinformation and democracy. Our goal here is to bring in the latest research to show what positions in the debate are valid and which could be discarded. In the second part we focus on the Flemish context and set out how these dynamics play out locally. In the report's third and final part we set out the current policy recommendations for tackling disinformation and democracy and then, in light of our understanding of the research and our review of the Flemish context, we explain the merits of limits of those recommendations. This allows us to conclude by introducing our own, evidence-led recommendations.

1. Disinformation & Democracy: A multi-layered debate

In the last years we have seen a growing interest in the speed by which disinformation has been spreading throughout the media landscape (Bradshaw & Howard, 2019). In 2018 83 percent of EU citizens felt disinformation was a danger to democracy (European Commission, 2018). People have been exposed to disinformation through social media, but also legacy media play a significant role as influencers in this exposure. Newspapers and broadcasters make these stories visible by linking to them, commenting on them or discussing them and thereby amplifying the stories. Topics such as climate change, refugees and immigration, and health-related issues such as vaccines and alternative medicine (e.g. during the Corona pandemic) are especially hot topics in disinformation campaigns. These stories often have a sensational character that compel legacy media to pick up on them. This expanding circulation of extremist opinions and interpretations has led to claims that democratic societies have become and are becoming more polarized in ways that damage democracy. Blame for the spread of extremist, polarizing content is often apportioned especially to social media and the algorithmic curation and moderation of content in general favoring especially powerful influencers of all kinds.

In Flanders, traces of this poisoning of the democratic well were already evident when we began this report in early 2019. The journalist Rudi Vranckx of public service broadcaster VRT NWS had to block accounts when he became the target for hate speech in 2018 (Vranckx, 2019). When we spoke to those involved in security and foreign policy, they noted that the presence of disinformation from agents *outside* Flanders and Belgium raised the question: *Why are Flanders and Belgium worth targeting*? One policymaker told us that it was because Flanders and Belgium stand strong for multilateralism, human rights, and a politics of compromise in the international sphere. They project a narrative that anti-liberal and anti-democratic forces are trying to undermine. In this way we see how even a smaller European country can become entwined in a broader geopolitical struggle. This highlights why an understanding of how disinformation becomes present in Flanders opens up our understanding of the mechanisms through which that geopolitical struggle is being waged.

Research across democracies shows that often the stories circulated take the form of hyper-partisan news rather than directly false information. This makes the widespread strategy of detection of 'fake' news weak and fragile and creates a problem for the protection of fundamental rights of freedom of expression across the political spectrum. In this section therefore, we will try to conceptualize the term disinformation, democracy and associated concepts and subsequently in the following sections discuss digital logics and existing studies and recommendations to combat the issues.

Conceptual definitions

The Council of Europe report (Wardle & Derakhshan, 2017) and the subsequent EU Commission report (Buning et. al., 2018) both focus on disinformation as a specific type of fake news that is defined as "false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit" (Buning et. al., 2018, p. 10). When policymakers and researchers conceptualize political fake news this is the definition most often used with associated potential threats to democratic political processes (e.g. protection of voters and free elections) and values (e.g. trustworthy information and transparency). In this way, we avoid talking about misinformation as information that is false or misleading but not necessarily intentionally designed to cause public harm. Here, we often find journalistic interpretations and reporting on research findings within for instance health studies (e.g. vaccines) that are controversial and conflicting with the majority of studies but support certain radical movements in society – and by being controversial also are shared among users due to the sensational character and amplified by bots that overflood the discourse and make it more divisive (Broniatowski et al., 2018; Nahon & Hemsley, 2013; Vosoughi et al., 2018).

The term fake news has been ascribed to the contemporary political scene and especially to the election campaign of Donald Trump (Billiet et al., 2019). This has created a renewed focus within academia on 'alternative facts', 'post truth', and on the philosophical problem of when something is true or false/fake. Even though most scholars agree that the ontological discussion on whether something is true or false depends on the person evaluating it (Farkas & Schou, 2019), other studies emphasize the politics of highlighting triangulation (Lyons, 2018) - and the search for truth with its web of interests unfolded - as an important democratic value that potentially can have self-censoring and adjusting behavior through social shaming designs. However, the effect of highlighting properly such verified stories within this heterogeneous, dynamic media environment is a contested topic favoring alternatives over flagging (Bode & Vraga, 2015) but further interdisciplinary research is needed in order to understand long-term effects (Lewandowsky et al., 2012; Tromble & McGregor, 2019).

People not only associate fake news with disinformation, misinformation, and 'fake' versus true information. Studies show that the concept is also associated with partisan political debate and 'poor journalism' more broadly, as well as being used to criticize news media and social media platforms (Buning et. al., 2018; Nielsen & Graves, 2018; Wardle & Derakhshan, 2017).

Deep fakes: Disinformation or an expression of evolving norms of communication?

A good example of the difficulty addressing dis- and misinformation is the case of deep fake videos. The videos are manipulated and thereby labeled deep fake, but they are partisan in the sense that it conveys a certain version of the truth -- they are made to portray the 'real' character flaws of those depicted. Critics would claim that the video format should assume a 1:1 representation of the real events and the claim of having this exact representation despite its manipulated character makes it fake, but this format expectation has changed over time e.g. through digital satire and memes. Also, docufiction existed before social media and did not offer an exact depiction yet was considered to convey some truth of events. Nevertheless, insisting on clear community rules for indicators of genre can be a way to steer user expectations and decrease the number of misleading videos as actors are not interested in disclosing the manipulated character of the video in propaganda and hyper-partisan content. Due to these logics a race towards automated fake detection will first of all be a race against time because actors will always try to circumvent community rules and algorithmic detection models and second it will contain some false positives that potentially will have damaging effects on societies and communities and the value of freedom of expression. An alternative would be to scale the number of people moderating (social) media content, and upgrading the moderation both in terms of education, salary construction (not pay per unit) and providing decentralized cultural moderation.

In this report we recognize the multi-faceted character of the topic and concepts (e.g. deep fakes, fake news, misinformation and disinformation), yet we recognize this under the notion of disinformation, drawing perspectives to other interpretations when it is relevant for the Flemish context of the report. Flanders is characterized as being a small country with high education levels and a multi-party system that influence challenges of disinformation differently to in the United States context, for instance. One of the challenges in applying the definition of disinformation from the EU Commission as "false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit" (Buning et. al., 2018, p. 10) is that it is very difficult to establish when something is intentional or not. This is because the network of actors involved in such a campaign can be vast and the stakeholders may be masked in more sophisticated campaigns using influencers, bots and third party 'sock puppets' to influence the frame and narrative of a discussion, the sentiment on a political topic, and/or the outcome of an online debate. In this report,

we instead understand disinformation from a more external conceptualization. We emphasize how disinformation is appearing and the impact it has had, not what the intentions behind its production and circulation may have been.

In this report we focus on the impact of disinformation on democracy. We understand democracy as a deliberative and participatory process of collective authorization and action. Parties and candidates offer manifestos which citizens can discuss in public and private before deciding for whom to vote. Through news media reporting and personal experience citizens can then follow and evaluate the elected party's performance, deliberate this, consider rivals' proposals, and then, at the next election, vote accordingly. Democracy is a looping circuit of accountability through time (Warner, 2003). It requires public and private spaces for informal discussion. Various institutions or 'organs' of democracy like parties, unions, community social hubs and news media enable and sustain those spaces of participation (Bryce, 1888). Those spaces collectively are often conceived as 'the public sphere' (Habermas, 1962). The quality of information citizens receive in that public sphere becomes critical since this is the basis on which deliberation, judgment and action rest (Dewey, 1927; Lippmann, 1921). Equally, even if citizens have equal formal voting rights, the ability of citizens to participate informally, freely and ---equally in the public sphere becomes a concern if social hierarchies and antagonisms lead to the formation of micro-spheres or "sphericules" (Gitlin, 1988) that mean citizens lack a common basis of information. Such micro-spheres may also diminish citizens' experience of and ability to understand and exchange views with people of different political positions.

The development of the digital realm has generated much research and understanding of digital spaces. Research has treated social media as a public sphere or as multiple public and semi-private spheres. Within those spaces or spheres, democratic ideals and values can be negotiated and different opinions meet in order to come to mutual agreements or just co-exist as a participatory democracy (Bechmann, 2019). On the one hand, social media allow for users to meet and discuss viewpoints and be exposed to issues and perspectives they had not considered before, potentially even drawing previously apolitical people to participate in politics (Vaccari, Chadwick, & O'Loughlin, 2015). On the other hand, studies show that self-selection of stories from people that we agree with is very high (Bakshy, Messing, & Adamic, 2015). In addition, the social media platforms' business model leads them in the design of algorithms, in the interface design and in the default settings to support this tendency in order to retain user time on the specific platform that then can be commodified in advertising revenue. This has led scholars to produce holistic accounts proposing human societies are inhabiting a new stage of politics, society and economy function largely, and in novel ways, through platforms (Couldry & Mejias, 2019; Pasquale, 2015; Zuboff, 2019). The issue in relation to democracy therefore becomes whether

platforms' algorithmic choices, interface designs, and default settings are an industry right or whether they are in conflict with democratic ideals and values. These democratic ideals and values include, for instance, the protection of voters' rights and transparency during elections – transparency about campaign narratives, actors and campaign budgets -- when there is high potential for their manipulation.

At the same time, it is important to note that these issues are not only isolated to specific election periods such as the recent 2019 European Parliament election, national or regional elections, or even to moments when a single issue is being debated in a parliament. These issues are also relevant in non-election periods where the sentiment around topics is building, where parties' performances are being evaluated, citizens informally deliberate their preferences, and policymakers may be sensitive to public opinion. At the same time, when we discuss the overall polarization in society, another way in which poor, non-triangulated journalism may have an effect on democracy concerns the potential echo chamber or filter bubble effects of self-selection and algorithmic curation in social media (Bechmann, 2017; Bechmann & Nielbo, 2018). However, to our knowledge we do not have any studies showing a direct effect between (social) media and an increase in polarization. On the contrary, we see no significant effects of political interest and echo chamber effects. Instead, people are exposed to a more diversified media landscape representing different opinions (Dubois & Blank, 2018; Dutton, Reisdorf, Dubois, & Blank, 2017; Nguyen and Vu, 2019) and there is no evidence they click on stories recommended to them by algorithms (Commissariaat Voor De Media, 2019). In the same way, studies that show effects of disinformation exposure on social media and voting behavior are sparse.

Dilemmas in understanding disinformation

- **What methods allow us to see disinformation?**

Policymakers and intelligence agencies seek real-time methods through which disinformation could be detected and its spread disrupted especially during short-term events such as elections and pandemics but in general we are far from having such tools. Recent years have seen some very compelling use of network analysis to map how disinformation spreads. However, labeling the information correctly is still a challenge and such analysis does not tell us how people interpret and understand those messages. After all, social media users regularly share content in order to mock it or support it, not because they believe it. These studies also are usually focused on a single medium where data are available, such as Twitter, and cannot tell us how disinformation is spread across platforms. Policymakers increasingly recognize that narratives are central to how people understand the world around them and often announce schemes for "countering" narratives, but no automated

system exists for detecting the structural features of narrative in text (scene, actors, dilemma, instruments, resolution). Verification of false or distorted images is a further technical problem. The state of the art, then, is a mixture of software-led network mapping of actors and disinformation alongside time consuming human-led, qualitative interpretation of what that disinformation means in a particular context.

Lack of access to social media data also prevents longitudinal and comparative research on how disinformation circulates in different democracies. Without such research it is very difficult to build valid explanations of why certain groups share certain content (Moeller & Bechmann, 2019). Thus, even if we had more advanced methodologies, lack of data would prevent the production of knowledge about disinformation.

- **How to coordinate the projection of a positive strategic narrative?**

To the extent that the EU and its member states could project a positive strategic narrative about their role in the world, about the viability of liberal democracy, and about how they will advance their values, coordination becomes a thorny problem. One policymaker pointed to the new EU rapid alert system, a chance to make stronger the interaction of EU and member-state levels and develop a holistic approach. It could have a proactive strategic content system and an alert system so countries can see what is happening in others and learn from how others respond effectively or not. Several journalists and policymakers named some EU countries such as the UK, Scandinavian states and some Baltic states as more advanced in their ability both to project a narrative proactively and to counter disinformation. While others such as Hungary were seemingly unaware of the multiple external sources of disinformation entering the EU and member states or ignoring it. A proactive system could be supported by the European External Action Service (EEAS). However, any such "sponsorship" would leave any narrative open to accusations of simply mirroring the state propaganda of Russia, China and others. In addition, institutional overlap (is disinformation a matter for defense, foreign ministries, police, intelligence, and so on) and serious political differences between member states make the formation of such a narrative difficult.

- **How much polarization is too much?**

How much polarisation is desirable in a democracy is not a given. In 1950 the American Political Science Association called for *greater* polarisation. In the post-1945 context, parties offered no clear, coherent differences of policy and ideology. This was felt to stymie voter choice and pluralism (Tucker et al., 2018). A range of types of filter bubble existing today, based on news source, algorithmic amplification, ideology, and social identity, each enabling different forms and degrees of polarisation. Is there a happy medium for a media system that would sustain agonistic democracy,

and how could that be achieved given that platform capitalism and populist politicians are incentivised to exacerbate difference, sensation and outrage? We will return to this in our recommendations.

- **Disinformation is not only on the internet**

It must be acknowledged that any solutions to the problems of disinformation that focus solely on online media will fail by default. Citizens inhabit a multi-media environment, one in which face-to-face conversations and traditional broadcasting still matter a great deal. For instance, research shows citizens in some countries still gain most exposure to politics through television (Ferrín et al., 2019). Disinformation may start with a false claim in a television discussion and then be picked up online and circulate, just as it may start with a deliberately false claim in an online forum that is picked up and reported by television or radio journalists.

This creates a cost problem. A genuine understanding of how disinformation gains traction in online *and* offline environments would be time consuming and extremely costly, involving ethnographic or anthropological research in political communities. All incentives lead researchers to study disinformation on Twitter because Twitter data can most easily be obtained. An understanding of disinformation based entirely on the role of disinformation on a single medium, a medium not even widely used in some countries, creates the danger that we have only an incomplete and skewed picture of disinformation.

- **How to disrupt spirals of disinformation?**

A key argument is that any solutions to disinformation in democracies must pay attention to cycles or spirals through which disinformation becomes normalized. This may include when leaders, journalists or citizens report on and thereby share extremist, divisive or false content, driving its re-circulation; or when social media platforms set their algorithms to amplify sensationalist content because this drives their commercial revenue (for further evidence of these spirals, see Dahlgren et al., 2019). These spirals are multi-stage processes with the possibility for intervention at multiple stages, but little incentive for any particular actor to take responsibility for the whole. For instance, social media companies present themselves as neutral intermediaries, not profit-driven publishers, and thereby eschew responsibility for content shared on their platforms and the social effects of that content.

There is a second dimension to these spirals. They involve an interaction between epistemic and identity politics. By epistemic politics we refer to the politics of what counts as true – as the processes and institutions through which a “fact” is agreed upon, whether in science, law, or journalism (Latour,

2018). By identity politics we refer to the politics of in-group and out-group: the power relations that determine who is seen as "us" and "them", what those boundaries mean, and how this bears on conduct. It may be that a fact-free media ecosystem makes it easier to demonise the "Other" and for the crimes of one's own leaders to be ignored, reinforcing an antagonistic identity group for those who feel they belong to that media ecosystem (Kreiss, 2019). Meanwhile, a fact-driven, public-service-oriented media ecosystem may affirm a very different relation to veracity and lead those who feel a belonging to this ecosystem to demean and ultimately antagonize those who fall outside this group. Divisions about the nature of who "we" are and the nature of what "we" believe can drive resentment in ways that reinforce and amplify division.

- **Are we at *war* to protect democracy?**

When disinformation is put within the context of politics and propaganda then it is framed as a matter of information warfare within democratic states waged by actors inside and outside those states. Here, media become "weapons", politics understood as war, and citizens must avoid becoming pawns of the enemy by sharing its disinformation (see e.g. *Scientific American*, 2019). Here, military and security organisations begin to understand the management of internet communication as an issue within their domain, with NATO for instance setting up STRATCOM centres in Eastern Europe for research on disinformation. There is extensive evidence that political campaign teams *have* tried to influence elections in other countries using techniques including micro-targeting (Jamieson, 2018; Joseff & Wooley, 2019) and the creation of pseudo-anonymous accounts (Friedberg and Donovan, 2019). These techniques may be legal or simply unregulated. They may comply with social media platforms' terms of service, or lie beyond the capacity of social media platforms to police. In 2017 Twitter found that nearly two-thirds of its users were displaying bot-like behaviour, removed them, but found they still could not keep up with the creation of bots (Allcott et al., 2019). These "weaponisation" techniques can be used to impersonate figures, to drive antagonism between identity groups, and undermine trust in leaders, journalists or citizens (Billiet, Opgenhaffen, Pattyn, & Van Aelst, 2019). This is one reason why many prominent intellectuals now voice fear for democratic culture – long-built, but perhaps quick to unravel (Habermas, 2016; Runciman, 2018; Wade and Mishra, 2018). The response of democracies has varied – leaders are unsure whether it is best to block external disinformation or trust citizens to understand what is disinformation and ignore it (Hellman and Wagnsson, 2017).

Russia -- and particularly the Putin regime – is also seen by some as an inspiration or model for political entrepreneurs and leaders within liberal democracies. The techniques used by the Putin regime to undermine trust in journalism as a system of delivering the truth of events for the Russian population serve as a model for others to win elections by undermining journalism and the

possibility of an open, fact-based public debate and instead focusing political discourse on identity (Pomerantsev, 2019). A study of 48 countries found that the use of computational propaganda occurred in all election campaigns (Bradshaw and Howard, 2018). Until this is regulated, why would any political actors *not* use these techniques given they are cheap and legal? It can be argued that, without effective regulation of such techniques, democracies have gone to war on themselves.

- **Is media literacy or fact-checking a solution?**

Can governments trust citizens to understand what is disinformation, and does fact-checking of news reports trigger citizens to reconsider their views of a story? Much research suggests this might not be straightforward. A study by Facebook found that when it flagged a news story as disputed and potential disinformation this did *not* make users less likely to click on it (Mena, 2019). Instead a "backfire effect" triggered some users to believe in the "facts" of the story even more (Smith et al., 2017). In the UK, a study of why users share news on social media during the 2017 General Election found that 16 percent of users knowingly share news stories they knew to be exaggerated or false, and this rose to 25 percent for those who share tabloid news every day. Given that tabloid news constitutes the most popular news sources in the UK, this indicates a widespread culture in which many citizens are comfortable sharing disinformation (Chadwick et al. 2018).

For this reason, many experts recommend treating disinformation as a collective social problem and not a matter of individual behaviour. Tufekci argues that social media had to be regulated as a public good and that legislation had to ensure data transparency at the individual level would generate solutions at the social level. She writes, 'Companies can and will persuade people to part with their data in ways that may seem to make sense at the individual level but that work at the aggregate level to create public harms' (Tufekci, 2018, no page). If users could see which advertisers were targeting them, how their data was being analysed and by whom, and if legislation ensured aggregated data was not used in discriminatory ways (for instance on health insurance) then this would produce a more open and data-literate culture (see also Buning et al., 2018). In short, the collective conditions for media literacy have to be created.

Fact-checking in mainstream media and on social media platforms could contribute to a collective solution because those verifications reach mass audiences. When we wrote this report there was already an international fact-checking network (IFCN) and experts we interviewed in Flanders discussed the creation of a coordination pan-European fact-checking network. However, one journalist in Brussels argued that professional journalism should already fact-check thoroughly and that organisations focused solely on fact-checking might miss the context or wider process within which a fact is placed. There are also questions about how shared norms of fact-checking could be agreed and whether EU-funding of a fact-checking network would politicise it.

We have set out how disinformation is defined, why it is significant for democracy, and the range of dilemmas faced by the full range of stakeholders involved – most importantly policymakers, journalists, social media companies, and citizens themselves.[1] In the next section we explore how these dynamics and dilemmas play out in a Flemish context.

[1] There are other stakeholders including advertisers, civil society organisations, and intelligence and police agencies who also have a direct interest and concern in the question of disinformation and democracy. However, given that this report analyses a range of policy and academic reports on the four stakeholder groups – policymakers, journalists, social media companies, and citizens – we have no scope to evaluate research and recommendations for additional stakeholder groups. We urge others to pick up this baton.

2. The Flemish context: Cracks in the system?

The picture of the Flemish context is that media and political institutions are widely trusted but may not be so resilient against an increase in disinformation. It is a small media market with sufficient subscription levels and public service broadcasting funding to ensure news organisations are not as financially vulnerable as in countries with larger and/or more commercial media systems. By the 2014 Belgian federal elections these established news organisations had achieved agenda-setting power online that reinforced the agendas they already generated in their daily broadcast and press outputs (Harder et al., 2017). By 2019, however, many journalists and policymakers we spoke to wondered whether this self-reinforcing position might be showing cracks – cracks that new or extremist news outlets could slip between and establish new foundations.

Belgium enjoys federal (national) and regional elections. Belgium features three regional substates: Flanders (Dutch-speaking), Wallonia (French-speaking) and the East Cantons (German-speaking). Each has equal autonomy and responsibility for matters of culture, education and health, and each has a parliament while also sending representatives to regional/municipal and federal governments and councils. Those overlapping bodies also offer different opportunities for parties to compete for the votes of the different communities based on language and identity considerations. For instance, in 2019 federal elections the extreme right-nationalist party Vlaams Belangs (VB, the "Flemish Interest") finished third nationally but second in Flanders. This intricate web of institutions and competitions has preserved a degree of peace and order for many decades.

But was this shifting? The political context drove the question forward. At the national level, Belgium's federal elections generally feature fierce competition between a large number of parties, who only require five percent in a constituency to qualify for seats. In May 2019 elections the secessionist New Flemish Alliance (N-VA) won with sixteen percent of the vote and took 25 out of 150 seats in the federal parliament. Strikingly, Vlaams Belang became third nationally, with 18 seats – something not predicted by any of the experts we met in April 2019. The potential for greater polarization is there because while consumers of subscription news may retain relative support for the traditional mainstream parties, there exists a section of the Flemish community who rely on free news content likely to be of lower quality and less likely to work to traditional journalistic norms of the mainstream public sphere. Local experts referred to 'alternative news' sites such as Sceptre.be and Doorbraak.be, whose numbers of readers or users may be low but whose content can be re-circulated via social media or taken up by political actors. .

Mainstream journalism in Flanders has investigated far right political groups (e.g. Verheyden, 2018). However, a danger is that if extremist groups receive a high vote share then current Flemish

journalism will find it difficult to deny them a platform. Research indicates Flemish journalists follow a hierarchy of power that produces a dynamic of 'centralized personalization' in Flanders that means that only a few political leaders receive a reasonable share of news coverage (Vos and Van Aelst, 2018; Wauters et al., 2018).

In Belgium as a whole, levels of trust in news media are among the top third among EU member states, especially printed and online newspapers and news magazines (European Commission, 2018). Media use and political satisfaction have grown in parallel in Belgium since the 1980s (Van Aelst, 2017). Across the EU, citizens of Belgium are least likely to say false news is a problem in their country (ibid.). Belgian citizens' encounters with news they think is false falls around halfway in the EU average. However, they express very low confidence in their own ability to identify false news and strongly believe it is up to media and political institutions to identify false news, not themselves (ibid.). This indicates a potential social fragility, as a trusting electorate may not be experienced in unpacking what makes a news report true or false.

We asked Flemish policymakers and journalists about fact-checking. Policymakers felt under a new level of pressure to verify information publicly. One gave an example concerning the Gillet Jaune protests in Paris in 2019. Journalists contacted Belgian policymakers to ask about a video of what appeared to be a Belgian husband allegedly crying "The French police shot my wife." It looked like fake news but the Minister could not risk immediately calling the story "fake news" because it was unclear if it was actually true. Hence policymakers now must perform verification and de-bunking work. They cannot disengage from this extra work. One journalist said that such unavoidable engagement created a dilemma. 'You have to engage with [the un-verified story], so immediately you recognise it as a "thing", and as soon as it's a "thing" you've got two separate arguments: which one is true?" Here we see how disinformation and the acceleration of responses driven by social media are creating dilemmas in the Flemish context that research has shown in other democracies.

However, policymakers we spoke to were confident that Flemish citizens were patient enough to wait for verification. As long as policymakers announced they were looking into an incident, they felt they were trusted to get the facts right. Against this, journalists are pressured to publish stories rapidly and cannot wait until their deadline passes, particularly as they know that the story is already on social media from other sources. Overall, we found policymakers felt pressured by journalists more than the Flemish public.

How does the dilemma of labeling disinformation part of "war" or the notion of a "war of narratives" play out in Flanders? We found little of such rhetoric. We did find that the Belgian Ministry of Foreign Affairs has digital diplomacy officers whose work relates to disinformation. Their role is: (i) projecting a narrative to external audiences and producing information for Belgian diplomats to promote Belgium; (ii) monitoring what is being said about Belgium in social and news media

overseas. This creates risks of blunders – young staff are given license to promote Belgium digitally and they are not tightly restricted, so the informal tone brings dangers of mis-representation. Nevertheless, this friendly register appears a calculated risk. The role of diplomats outside of Flanders could be useful too, one interviewee told us. A core job of diplomats posted to embassies overseas is to identify what is *real*; to notice influences in a territory. They become aware of the conspiracy theories and narratives of that country. It would seem a natural continuity to take advantage of diplomats who are aware of, and can be used to work against, disinformation.

What we find in Flanders reflects what studies of public diplomacy and digital diplomacy have found more widely: that social media both creates challenges for democratic countries' diplomatic services -- to be always monitoring, always-on, and skilled in digital communication -- but also opportunities to spot negative disinformation early, to positively manage national image, and to coordinate narratives with allies.

3. A turn in the debate: Existing recommendations and our proposals

Evaluation of existing recommendations

There have been several reports and white papers on disinformation in Europe both within nation states and at European Union level. Many reports refer to the EU Commission report on disinformation (Buning et al., 2018) and the recommendations have directly been the background for many actions taken from the European Commission against disinformation such as the European action Plan and Code of Practice:
https://eeas.europa.eu/sites/eeas/files/action_plan_against_disinformation.pdf;
https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation.

This EU Commission report lists five recommendations that we will discuss in a research perspective in this section, leading up to a discussion on our research-based recommendations, focusing on the Flemish context.

Enhance transparency of the digital information ecosystem

In the execution of the European Action Plan, transparency has been one of the challenges to implement in practice. This is despite the platforms signing an agreement to support this through increased data sharing (see also Moeller & Bechmann, 2019). The main reason for this according to platforms is that transparency risks violating user privacy. According to our interviews and meetings the platforms refer to GDPR as a large barrier to the ability to act on transparent solutions for scrutiny for independent researchers and journalists. This concern has accelerated in the light of Cambridge Analytica using Facebook data gathered for research in political campaigns using micro-targeting. On the other hand, European politicians refer in our meetings to GDPR as very liberal when it comes to academic research and therefore GDRP *cannot* be used as an argument against sharing data in safe space solutions. The European Action Plan builds on good faith research "soft law" solutions. In one solution, for instance, all platforms have made an Ad Transparency Library or Report in order to create transparency about which ads have been exposed on the platforms, who is the source, how many of each ad was posted, and who on a very general level (age, region) has been exposed to the ad in question. These archives are positive results of the Action Plan and also create possibilities for small regions like Flanders to investigate political ads. Yet, the Ad Transparency Libraries follow different data formats across platforms – some with application programming interfaces (APIs) like Facebook and some without like Google. They are not designed to allow large-scale comprehensive analysis (data science) and do not allow for researchers and journalists to data mine the text and pictures of the ads in order to understand the potential microtargeted variance (Moeller & Bechmann, 2019). As we have seen too in the recent UK 2019

General Election in which Facebook's campaign ad library mysteriously deleted major parties' ads 24 hours before polling day (Manthorpe, 2019), these systems are unreliable.

Another solution for research has been Social Science One, a collaboration between social science research council (SSRC) in the U.S. and Facebook (King & Persily, 2018; see more details in Moeller & Bechmann, 2019). Here, Facebook is providing access to only selected data grant owners to execute projects in a protected environment without data exchange taking place and with a privacy account that only allows for a certain number of data points to be modulated in the research in what is called a "differential privacy solution" (Dwork, 2008). This disfavours small countries such as the Flemish region of Belgium as it is easier to disclose identity in such countries due to the population size (Moeller & Bechmann, 2019). Also, journalists can apply for access to open data through Social Science One using the module CrowdTangle (Open groups and pages). However, at the time of writing, Social Science One has been highly criticized for not providing successful access and the funding bodies behind Social Science One recommended "winding down the project" since Facebook has been unable to deliver (Silverman, 2019). This further accelerated in December 2019 where the European representatives along with co-founders Gary King & Nathaniel Persily officially criticized Facebook for not sharing the data agreed upon and thereby openly questioned the construct (Vreese et al., 2019).

Solutions such as Social Science One and Twitter data grants can be seen as problematic for the research community broadly. Access is only granted to a selected few, potentially favouring American and well-esteemed labs/universities and thereby destroying the community's ability to assess this type of research and create critical mass to make peer oversights. Also, the character of the "solution" may undermine academic freedom because it only allows for access to specific types of datasets designed to answer specific questions, instead of sharing data that comply with European regulation and is overseen by national and university-based data agencies and international review boards. Furthermore, due to privacy being protected strongly on both the actor and the data unit level, actual content analysis becomes highly problematic as this type of content is made to disclose user identity. This problem is increasing in end-to-end encrypted platforms such as WhatsApp. However, these platforms are also communication spaces in which various forms of disinformation is circulated. Certainly, researchers should not have access to private conversations. But from a societal point of view it is essential that researchers under the right safe space solutions are able to access and analyse communities of a certain critical mass in order to safeguard democracy and democratic values.

Promote media and information literacy

The report from the EU Commission also highlights more media and information literacy for users as well as key stakeholders such as influencers, journalists and politicians. An increased evidence-based approach to knowledge of online information disorders is something that everyone supports. Yet, the discussion point is if we are looking at the problem from the right angle – are we asking the right questions?

Up until now the major focus in the discussion on information disorder has been on the content, whether something is true or not, harmful or not and should be taken down in order to secure democratic values. However, this practice has led to claims of violation of freedom of expression and in a Deweyian sense actually prevent different voices to meet and negotiate value of the specific society. On the other hand, we know from previous studies that the platforms amplify sensational content (Nahon & Hemsley, 2013) and that disputed content by definition is sensational.

In the same way we have too little research that actually show that fact-checking has an effect on the content circulation even though it might have a political signal value (Mena, 2019, Bode & Vraga, 2015). Actually, we have studies outside Belgium pointing to users being in a completely other state of mind than the rational one when they share news, that it does not really matter whether it is true or false as long as it supports their worldview (Nielsen & Graves, 2018). Another study from the U.S. show that on Twitter 0.1 % of users accounted for 80% of fake news sources shared and those who were most likely to engage with such content were conservative, older and highly engaged with political news (Grinberg et al., 2019). Belgium users still rely heavily on traditional news and therefore these tendencies may be less likely but we still need studies that investigates this.

A study from U.S. shows specifically older people over 65 share more disinformation (Guess, Nagler, & Tucker, 2019) so any user-directed media literacy initiatives would have to target this age group. However, more studies are needed in order to better understand non-U.S. contexts along with the psychological incentives, motives and needs in order to improve literacy initiative and to measure possible effects before rolling out large literacy infrastructures.

Reducing disinformation to the content level by detecting and deleting content potentially will both violate freedom of expression and have limited effect as it will always be a race against time due to the global digital media networks where content is traveling fast and where new content in new formats from new sources is uploaded constantly.

Following these research contributions there might be an interest in shifting the focus on media literacy away from fact-checking and onto questions about how content is circulated, how content is being amplified, and what at a given time are useful indicators of harmful source characteristics and other logics for harmful actors. Yet, here again data access is a prerequisite for evidence-based approaches to information, data and media literacy as a potential power for social

change. We still need more experimental setups in order to understand effects of such literacy programs if the goal is that we stop sharing, liking and commenting and otherwise circulating disputed and debunked content. Again, recall that we have no studies showing large effects on voting behavior from having been exposed to disinformation.

Another blind spot in terms of information and media literacy is the need to investigate further effects of various “forced” interventions into the algorithm. For instance, what is the effect of ranking some content types over others. For instance, reducing click bait and sources with previous flagged or debunked content. Providing opposing views to disputed content has in existing experiments carried out by Facebook shown to be difficult for people to click on because we interact with what we agree on (confirmation bias) making social media opinionated spaces of interaction (Mena, 2019).

Develop tool to enhance the empowerment of users and journalists and foster positive engagement

The recommendation of tools to enhance the empowerment of users and journalist in the context of disinformation and foster positive engagement is an interesting point that ideally will build on evidence-based knowledge of strategic narratives and collective behaviour logics and especially on social media. However, as pointed out earlier existing studies (e.g. Mena, 2019) point to such use primarily being directed to influencers, journalist, and other bridging hubs even though strong ties are central to the circulation of content on social media (Centola, 2018) as flagging debunked content on the user interface of end-users will potentially have the opposite effect as it will then be sensational and create virality (Mena, 2019; Nahon & Hemsley, 2013; Lyon, 2017; Bode & Vraga, 2015). However, we do not know whether this is the case in the Flemish context as more studies are needed outside the U.S.

The natural language processing (NLP) research community has made excellent progress in detecting disinformation on social media (Aldwairi & Alwahedi, 2018; Lozhnikov, Derczynski, & Mazzara, 2018; Monti, Frasca, Eynard, Mannion, & Bronstein, 2019; Shu, Sliva, Wang, Tang, & Liu, 2017; Tacchini, Ballarin, Della Vedova, Moret, & de Alfaro, 2017) and manual factcheckers (here especially journalists) have made progress by uniting themselves on an international level through the international fact-checking network (IFCN). Also, platforms are buying manual fact-checking from journalists in order to train their machine learning algorithm to be better at detecting disinformation. However, both manual and automated fact-checking is a moving target due to socio-technical developments. In order for IFCN, factcheck.org and European equivalents to optimize the initiative for improved crowdsourcing more work needs to be done in order to standardize open source data models for distributed fact-checking and subsequent academic research. Such lists will also make

research on debunked content more precise as existing research primarily is based on source debunking, not story debunking (e.g. Vosoughi, Roy, & Aral, 2018). Another missing link in the tools available now (see also https://www.disinfobservatory.org/the-observatory/) is a stronger focus on positive engagement that cuts across detecting disinformation. This for instance, could be materialized as solutions for decreasing hostile sentiments or the use of clickbait (Kumar et al., 2018).

Safeguard the diversity and sustainability of the European news media ecosystem

At the moment news media are powerful amplifiers of disinformation and therefore it is central to look at the economic sustainability of this new ecosystem and to preserve diversity in order not only to foster (hyper)partisan and opinionated news stories online that make use of clickbait to go viral and drive traffic up (Bechmann, 2017). There has been a steep decline in revenue streams going from analogue to web to mobile and therefore news media in general have difficulties keeping a solid revenue and this may lead to alternative uses of amplifying logics that does not correspond to journalistic ideals of fact-checking such as for instance objectivity as source triangulation (Harcup & O'Neill, 2017). More research work and experiments need to be done in order to test effects of various models in order to encourage such ideals and to test whether in fact we see a strong correlation between business model (e.g. state supported vs. non-state supported) and the use of for instance clickbait. In the same way, there are no real alternatives to dominating platforms such as Facebook and Instagram that works under pure commercial incentives and through American ideals. Europe has a weak history in creating tech alternatives that are not acquired by American companies whereas China has had more successes (e.g. TikTok – https://www.tiktok.com/en/). More research is needed in order to understand how we can build a more sustainable system for growing European tech-alternatives that build on European ideals of safeguarding democratic values that has been the core of the strong public service traditions that exists in the Flemish context as well as elsewhere in Europe, such as the U.K., Scandinavia and Germany. Another strand of missing research is to investigate to what extent existing American platforms can have obligations towards European democracies that does not only build on 'soft-law' solutions as per the existing initiatives the EU Commission has applied.

Continuous research on the impact of disinformation in Europe

As outlined, the above recommendations cannot be executed unless we have continuous research on the topic of disinformation and its impact on democracies. Yet, this evaluation has also made clear that executing such research right now is in a fragile state as social media data access is at best problematic at worst non-existing (Moeller & Bechmann, 2019). And access to good (!) data is

a prerequisite for evidence-based research that can enhance our knowledge on the field and make fertile policy. Proper (and relatively fast) access is especially important in times of crisis if researchers should help understand logics and mitigate against damage to democracies e.g. in elections or under pandemics.

Prioritizing academic research is a way to mitigate between very strong stakeholder incentives. Set aside good faith and great individuals, the infrastructure is designed for optimizing certain incentives. A cynical analysis of motives and incentives would be that platforms are interested in good press and to keep stock holders happy, politicians are interested in elections and having channels, platforms and a media landscape that are suitable for their political messages, academics are driven by providing the next big research contribution by gaining exclusive data access, journalists are interested in getting back to being the centre of the news ecosystem in terms of attention and/or revenue, and users are primarily focused on their social positioning and recognition, own motives and needs in their daily use. Acknowledging that academics also have interests that do not necessarily serve the greater good is important in order to pursue an evidence-based knowledge approach that are sustainable in the longer run. Here, it becomes important to emphasize solutions on, for instance, data access that can scale across research communities and is not only being granted to the select few (Moeller & Bechmann, 2019).

Final remark: Contesting Narratives about Disinformation

There is a need in liberal democracies, including Flanders and Belgium, for attention on positive and productive relations in society and politics as well as negative and disruptive relations. The very idea of "democracy in crisis" *has effects*. If citizens are told that trust in political institutions is collapsing, that politics has been taken over by extremists and that social media is clogged with false information and dummy accounts then they will be more inclined to believe this is true and disengage further. Words in politics can become self-fulfilling prophecies. This is why illiberal forces promote the narrative of democracy in crisis.

This is not to discount the existence of disinformation or the economic, political and cultural conditions that enable it. Nor is this to distract from the more technical need we described above for a stronger focus on finding positive engagement online that cuts across information bubbles and undermines hostile content and intentions. Instead, we urge the adoption of a particular tone towards disinformation and for leaders – in politics, media, commerce and in communities – to direct attention to when democracy is working well and when extremist groups lose or wither. This is difficult. Bad news sells. Failing institutions become a problem that politicians can promise to solve. Yet it is possible. During the decades of conflict in Northern Ireland in the late Twentieth Century journalists and many politicians realized that directing attention to violence only drove further cycles of revenge

and outrage. “Sensationalism costs lives” became a maxim, and editors across political and religious lines agreed to avoid sensationalist reporting of terrorist attacks. Such action was counter to economic incentives, but it showed collective responsibility could be taken because of a common interest. Just as peace outranks one’s “side” winning, so democracy is more important than online clicks and buzz. We urge, therefore, the need for reflection, balance, and a commitment to social responsibility.

Concluding recommendations and future directions

Based on our conceptual discussions, accounts of the Flemish contexts and the discussion of existing recommendation and research we propose the following concluding recommendation and future directions for some of the key stakeholders. Please note that these recommendations would create a positive spill-over of increasing the knowledge in society about data, media and democracy and thus improve the possibility of evidence-based media literacy.

Political leaders:

- Short-term:
    - Independent (high level) statement from the EU Commission announcing that data exchange does not violate privacy and GDPR
    - Political parties must agree to not use bots, sock puppets or other third-party techniques that hide identity in order to spread content
    - Political parties must continue to agree to and expand the extent of disclosed budgets, commercial partnerships and content of political campaigns
    - Support building debunked lists at a story level in an open source solution with international data model format that encourage crowdsourcing
- Medium-term:
    - Political leaders must seek consensus on how to understand the tradeoff between privacy and transparency. They must be willing to enforce this and support this balance to external actors (the US, China, and so on) because this is a global phenomenon transgressing national borders.
    - Law and actions taken in Europe must be based on EU regulations, not US regulations.
- Long-term:
    - The EU Commission can provide non-financial support to establishment of a data exchange solution where data is stored outside and across platforms. Further work needs to be done in order to find optimal safe space solutions like DNA registers and

other health data, and to create practices for private company data to be stored this way.

- Commitment to disclose the use of social media fully by political actors both financially and in terms of narratives and content in order to prevent divisive strategies and hostile sentiments
- Politicians should create stronger support for research in how we create a democratic sustainable tech and media landscape in Europe that can provide alternatives to platforms operating under American values and commercial incentives. Such alternatives would ideally operate in parallel with the continuation of strong press and public service traditions that favour democratic values.

Journalists:

- Short-term:
  - Continue to 'call out' journalists who reprint disinformation, either by other journalists/editors (as a new norm) or through industry bodies.
- Medium-term:
  - Educate journalists in how platforms work (e.g. through Flanders' universities' masters degrees in Journalism), how data science works, and thus how to produce scientifically-informed reporting about platforms and using platform data.
- Long-term:
  - Build a culture of responsibility in which drawing attention to disinformation is done cautiously and reproducing disinformation is a severe norm violation (Model: In Northern Ireland during the conflict journalists on all sides agreed not to report violence sometimes, because such reporting just encouraged revenge attacks and more violence. The journalist norm was "sensationalism costs lives' so they collectively agreed to de-sensationalise their news). This would counter the 'escalating' and 'normalising' dynamic that journalist reporting of radical/extreme views has, making them mainstream and just encouraging even more extremism.

Platform companies:

- Short-term:
  - Develop high ethical standards applying to the European context in collaboration with European research communities

    - Making data available in a format that allows for academic researchers to pose own research questions in the belief of academic freedom and facilitate the independent investigation of potential information disorder in a scalable data access solution
    - Making deep link lists of debunked and flagged stories available for independent research through scalable safe space solutions that consider the privacy of data subjects and the societal interest of disclosing democratic deficiencies
    - Prioritize manual Flemish-based content moderation in order to control the circulation of disinformation in the form of divisive content using misinformation and hostile sentiment
    - Prioritize the downgrading of hostile sentiment and containment of hyper-partisan content in the communities they were intended for by monitoring and setting up literacy programs for bridging hubs/influencers/legacy media together with support for independent experimental research following strict European legal and ethical standards
- Medium-term:
    - Provide specific and differentiated APIs for (i) journalists and NGOs, and (ii) university-based academics in order to meet different GDPR concerns and societal interests
- Long-term:
    - Provide researchers with stable access to social media data in safe, controlled spaces that reduce privacy risks to users whose data is analysed, preferably outside platforms and across platforms.
    - Making internal research with a focus on to what extent social media data can be made non-identifiable from the beginning without compromising the need for opening up to academic researchers to pose research questions in the interest of society and with legal and ethical clearance

Academic researchers:

- Short-term:
    - We propose to adapt a conceptualization of (mis- and) disinformation that is broader than the EU Commission focus on “intentional” harmful content as intentions are hard to define and the problematic effect of misinformation can be harmful in a longer time span in terms of general level of trust and polarization in our society. These effects need to be further scrutinized

    - The academic community needs to push politically for access to good data from platforms in order to carry out evidence-based research at scale
    - The academic community needs to push for data exchange solutions that acknowledge the need for critical questions and a variety of methods such as qualitative, quantitative and experimental techniques in the support of academic and scientific freedom for the benefit of society
    - Academic communities and national foundations are encouraged to support and encourage this topic to be conducted collaborating across fields and across countries in order to recognize this as an interdisciplinary and international challenge
- Medium-term:
    - More research is needed in order to move away from a content-centric research agenda in the context of disinformation and instead move towards circulation at scale and across platforms
    - More experimental research is needed in order to measure effects of fact-checking and media literacy initiatives, reconfigurations of algorithms as well as networks, in order to slow down or prevent circulation of disinformation and hostile sentiments
    - More work is needed in order to understand effects between disinformation exposure and voting behavior
    - More research is needed in order to connect research on disinformation with larger questions of polarization and trust, as such causal connections are sparse in existing studies – we need more studies to back this up, and especially outside the UK and US, in small countries with high education level such as the Flemish region.
- Long-term:
    - Negotiating solutions for data access, processing and storage that cut across individual academic incentives for the benefit of the general knowledge level and skillsets in our research communities for researchers at all levels both within small countries and larger regions

4. Conclusion

This thinkers' report has shown how more research is needed in order to fully account for the potential consequences of disinformation sharing for the functioning of democracy although it has pointed to existing studies outlining among others issues of flagging, self-selection, incentives and motives anchored in users' own opinions. In continuation of the existing studies and the suggested recommendations this thinkers' report has also argued for a turn in the debate on democracy and disinformation. This turn is visible on different levels.

First, the conceptual debate on disinformation was initially laid out in the European Council and European Commission as an alternative concept to 'fake news' constructed during the Donald Trump campaign for political purposes. However, the recent discourse around disinformation takes a turn in order to include more elements than "false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit" to understand the logics of information disorders. The construction of alternative discourses are still key elements in disinformation, but at the same time includes other elements disregarding intent. Such elements are for instance misinformation, hostile sentiments and hyper-partisan information. Therefore, this report argues for a broader notion of disinformation to take such logics into account disregarding intentionality.

Second, we argue for a turn in the way access to data is provided by platforms to independent academic investigators. This access was provided as an open access for business, but in some cases a complete shutdown occurred resulting in effectively closing off much social media data from critical scrutiny by academic researchers. Now a few researchers have been granted access through data grants but this model relies on an idea that platforms will be able to anonymize data units, which has proven difficult if not impossible. We argue for a turn in this approach and instead of protecting data subjects on the data unit level we argue for a protection of the data subject through a strict access control and safe space solutions. We know them from DNA registers and census data registers that independent academic researchers have a long and strong experience in interfacing with in western communities such as the Flemish community. We recognize the need for democracies to have access to independent scrutiny in order to safeguard scientific freedom and in order to build strong evidence-based policies.

Third, the report has argued for a turn in strategies applied to counter disinformation on a strategic level. We argue for turning the attention away from only focusing on content and whether this is fake or debunked – and thus important to take down either through automatic detection or manual moderation – to a stronger focus on monitoring, analysing and containing amplifying effects and circulation logics of social and legacy media in digital ecosystems. This will also create a more

fruitful discussion on the matter of basic human rights such as freedom of expression. This is important to preserve for the democratic discussion. This turn has at least two implications. It forces us to discuss media literacy and fact-checking on two different things and with different potential interventions: as something directed towards end-users and factchecking initiatives directed towards influencers and journalists. It also forces demands on stakeholders, such as politicians, platforms and news media, to focus more holistically and not only see this as an issue of detecting false content that can automatically be taken down. The way forward is instead to fertilize discussions about democracy-friendly constructions of digital business models, discussions that address questions of advertising and other data-related economies. In this respect media literacy must be understood as a product of deep understanding of digital infrastructure logics as well as individual and collective behavior, rather than a technical matter of detection and take down procedures. For as we have argued in this report, the collective conditions for media literacy have to be created.

5. References

Aldwairi, M., & Alwahedi, A. (2018). Detecting Fake News in Social Media Networks. *Procedia Computer Science*, *141*, 215–222. https://doi.org/10.1016/j.procs.2018.10.171

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics*, *6*(2), 1-8. DOI: 2053168019848554

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Bechmann, A. (2017). The Facebook newsfeed as a news source. *Papers from Nordmedia 2017*. Presented at the Nordmedia 2017, University of Tampere, Tampere, Finland. Retrieved from http://pure.au.dk/portal/en/persons/anja-bechmann(b093426c-3157-465e-bd29-fe73ee20d446)/publications/the-facebook-newsfeed-as-a-news-source(a4f1cfdd-b2f4-4800-b013-2a32c08e1a59).html

Bechmann, A. (2019). Data as Humans: Representation, Accountability, and Equality in Big Data. In (ed. Jørgensen, R. F.) *Human Rights in the Age of Platforms,* Information Policy Series: Vol. 16. (pp. 73–94). Cambridge, MA: MIT Press.

Bechmann, A., & Nielbo, K. L. (2018). Are We Exposed to the Same "News" in the News Feed? *Digital Journalism*, *6*(8), 990–1002. https://doi.org/10.1080/21670811.2018.1510741

Billiet, J., Opgenhaffen, M., Pattyn, B., & Van Aelst, P. (2019). *The battle for the truth. Fake news and disinformation in the digital media world* (KVAB POSITION PAPERS No. 62b). Brussels: The Royal Flemish Academy of Belgium for Science and the Arts.

Bode, L., & Vraga, E. K. (2015). In Related News, That Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media. *Journal of Communication*, *65*(4), 619–638. https://doi.org/10.1111/jcom.12166 https://doi.org/10.1126/science.aau2706

Bradshaw, S. & Howard, P.N. (2019). *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*. Oxford: Oxford Internet Institute. Available at: https://comprop.oii.ox.ac.uk/research/cybertroops2019/

Bradshaw, S. & Howard, P.N. (2018). *Challenging Truth and Trust: A Global Inventory of Organised Social Media Manipulation*. Oxford: Oxford Internet Institute. Available at: https://comprop.oii.ox.ac.uk/research/cybertroops2018/

Brehm, J. W. (1966). *A theory of psychological reactance*. Oxford, England: Academic Press.

Brussels Times, The (2019). Vlaams Belang spent € 3.2 million to finance election campaign. 22 August. Available at: https://www.brusselstimes.com/all-news/belgium-all-news/64843/vlaams-belang-spent-e-3-2-million-to-finance-mays-eu-campaign/

Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., … Dredze, M. (2018). Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. *American Journal of Public Health*, *108*(10), 1378–1384. https://doi.org/10.2105/AJPH.2018.304567

Bryce, J. B. (1888). *The American commonwealth*. London: Macmillan.

Buning et. al., M. d. C. (2018). *A Multi-dimensional Approach to Disinformation: Report of the independent high level group on fake news and online disinformation* (pp. 1–40). Retrieved from EU Commission website: https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation

Centola, D. (2018). *How Behavior Spreads: The Science of Complex Contagions*. Princeton, NJ: Princeton University Press.

Commisariat Voor De Media (2019). Filter Bubbles in the Netherlands. 20 May. Available at: https://www.cvdm.nl/wp-content/uploads/2019/06/publication-Filter-bubbles-in-the-Netherlands.pdf

Dahlgren, P. M., Shehata, A., & Strömbäck, J. (2019). Reinforcing spirals at work? Mutual influences between selective news exposure and ideological leaning. *European Journal of Communication*, *34*(2), 159-174.

Couldry, N., & Mejías, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford, CA: Stanford University Press.

Dewey, J. (1927). *The public and its problems: An essay in political inquiry*. New York.

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, *21*(5), 729–745. https://doi.org/10.1080/1369118X.2018.1428656

Dutton, W. H., Reisdorf, B., Dubois, E., & Blank, G. (2017). *Search and Politics: The Uses and Impacts of Search in Britain, France, Germany, Italy, Poland, Spain, and the United States* (SSRN Scholarly Paper No. ID 2960697). Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=2960697

European Commission (2018). Flash Eurobarometer 464, February. Available at: https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/survey/getsurveydetail/instruments/flash/surveyky/2183

Farkas, J., & Schou, J. (2019). *Post-Truth, Fake News and Democracy: Mapping the Politics of Falsehood*. Routledge.

Ferrín, M., Fraile, M., & García-Albacete, G.M. (2019). Who Learns in Information Rich Contexts? The Informative Effects of the 2015 Spanish Electoral Campaign. *International Journal of Press/Politics*. Online first, DOI: doi.org/10.1177/1940161219832455

Friedberg, B. & Donovan, J. (2019). On the Internet, Nobody Knows You're a Bot: Pseudoanonymous Influence Operations and Networked Social Movements. *Journal of Design and Science*. Issue 6, 7 August. DOI: 10.21428/7808da6b.45957184

Gitlin, T. (1998). 'Public Sphere or Public Sphericules?' in Liebes, T., Curran, J. (eds), Media, Ritual and Identity, Routledge, London, pp. 175–202.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, *363*(6425), 374–378. https://doi.org/10.1126/science.aau2706

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), eaau4586. https://doi.org/10.1126/sciadv.aau4586

Habermas, J. (1962). *The Structural Transformation of the Public Sphere: An Enquiry into a Category of Bourgeois Society*. Cambridge, MA: The MIT Press.

Harcup, T., & O'Neill, D. (2017). What is News?: News values revisited (again). *Journalism Studies*, *18*(12), 1470–1488.

Harder, R. A., Sevenans, J., & Van Aelst, P. (2017). Intermedia agenda setting in the social media age: How traditional players dominate the news agenda in election times. *The International Journal of Press/Politics*, *22*(3), 275-293.

Joseff, K. & Woolley, S. (2019). *The Human Consequences of Computational Propaganda: Eight Case Studies from the 2019 U.S. Mid-Term Elections*. Available at: https://www.iftf.org/disinfoeffects

Kreiss, D. (2019). From Epistemic to Identity Crisis: Perspectives on the 2016 US Presidential Election. *International Journal of Press/Politics*. 24(3), 383-388.

Kumar, V., Khattar, D., Gairola, S., Kumar Lal, Y., & Varma, V. (2018). Identifying Clickbait: A Multi-Strategy Approach Using Neural Networks. *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 1225–1228. https://doi.org/10.1145/3209978.3210144

Latour, B. (2018). *Down to Earth: Politics in the New Climatic Regime*. Cambridge: Polity. Trans. Catherine Porter.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*, *13*(3). https://doi.org/10.1177/1529100612451018

Lippmann, W. (1921). *Public Opinion.* New York: Macmillan.

Lozhnikov, N., Derczynski, L., & Mazzara, M. (2018). Stance Prediction for Russian: Data and Analysis. *ArXiv:1809.01574 [Cs]*. https://doi.org/10.13140/RG.2.2.15252.76161/1

Lyons, B. A. (2018). When Readers Believe Journalists: Effects of Adjudication in Varied Dispute Contexts. *International Journal of Public Opinion Research*, *30*(4), 583–606. https://doi.org/10.1093/ijpor/edx013

Manthorpe, R. (2019). Researchers fear 'catastrophe' as politics ads 'disappear' from Facebook library. Sky News, 11 December. Available at: https://news.sky.com/story/researchers-fear-catastrophe-as-political-ads-disappear-from-facebook-library-11882988

Mena, P. (2019). Cleaning Up Social Media: The Effect of Warning Labels on Likelihood of Sharing False News on Facebook. *Policy & Internet*, *online first*. https://doi.org/10.1002/poi3.214

Moeller, L. A., & Bechmann, A. (2019). *Research Data Exchange Solution* [Report for the EU Commission]. Brussels: SOMA.

Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M. M. (2019). Fake News Detection on Social Media using Geometric Deep Learning. *ArXiv:1902.06673 [Cs, Stat]*. Retrieved from http://arxiv.org/abs/1902.06673

Nahon, K., & Hemsley, J. (2013). *Going viral*. Cambridge, MA: Polity Press.

Ngyuen, A. & Vu, H.T. (2019). Testing popular news discourse on the "echo chamber" effect: Does political polarisation occur among those relying on social media as their primary news source? *First Monday*. 24: 6. Available at: https://firstmonday.org/ojs/index.php/fm/article/view/9632

Nielsen, R. K., & Graves, L. (2018). *"News you don't believe": Audience perspectives on fake news | Reuters Institute for the Study of Journalism*. Retrieved from https://reutersinstitute.politics.ox.ac.uk/our-research/news-you-dont-believe-audience-perspectives-fake-news

Pasquale, F. (2015). *The black box society*. Cambridge, MA: Harvard University Press.

Pomerantsev, P. (2019). *This is Not Propaganda: Adventures in the War Against Reality*. London: Faber and Faber.

Scientific American (2019). Everyone Is an Agent in the New Information Warfare. September. Available at: https://www.scientificamerican.com/article/everyone-is-an-agent-in-the-new-information-warfare/

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ArXiv:1708.01967 [Cs]*. Retrieved from http://arxiv.org/abs/1708.01967

Smith, J., Jackson, G., and Raj, S. (2017). Designing against misinformation. *Medium*, 20 December. Available at: https://medium.com/facebook-design/designing-against-misinformation-e5846b3aa1e2

Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & de Alfaro, L. (2017). Some Like it Hoax: Automated Fake News Detection in Social Networks. *ArXiv:1704.07506 [Cs]*. Retrieved from http://arxiv.org/abs/1704.07506

Tromble, R., & McGregor, S. C. (2019). You Break It, You Buy It: The Naiveté of Social Engineering in Tech – And How to Fix It. *Political Communication*, *36*(2), 324–332. https://doi.org/10.1080/10584609.2019.1609860

Tucker, J.A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D, & Nyhan, B. (2018). Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. Hewlett Foundation, March. Available at: https://hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf

Tufekci, Z. (2018). We Already Know How to Protect Ourselves From Facebook. *New York Times*, 9 April. Available at: https://www.nytimes.com/2018/04/09/opinion/zuckerberg-testify-congress.html

Vaccari, C., Chadwick, A., & O'Loughlin, B. (2015). Dual screening the political: Media events, Van Aelst, P. (2017). Media Malaise and the Decline of Legitimacy. In van Ham, C., Thomassen, J. J., Aarts, K., & Andeweg, R. B. (Eds.) *Myth and reality of the legitimacy crisis: explaining trends and cross-national differences in established democracies*. Oxford: Oxford University Press, 95-114.

Verheyden, T. (2018). VRT probe reveals racism and sexism at heart of rightist organisation, VRT NWS, 10 October. Available at: https://www.vrt.be/vrtnws/en/2018/10/10/vrt-probe-reveals-true-face-of-rightist-organisation/

Vos, D., & Van Aelst, P. (2018). Does the political system determine media visibility of politicians? a comparative analysis of political functions in the news in sixteen countries. *Political Communication*, *35*(3), 371-392.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Vranckx, R. (2019). “Dit jaar al veelvoud aan Twitter-haat”: Rudi Vranckx blokkeert voor het eerst accounts. *De Morgen*, 10 January. Available at: https://www.demorgen.be/nieuws/dit-jaar-al-veelvoud-aan-twitter-haat-rudi-vranckx-blokkeert-voor-het-eerst-accounts~b3f4c8ce/

Vreese, C. de, Bastos, M., Esser, F., Giglietto, F., Lecleher, S., Pfetsch, B., … Persily, N. (2019). Public statement from the Co-Chairs and European Advisory Committee of Social Science One. Retrieved December 15, 2019, from https://socialscience.one/blog/public-statement-european-advisory-committee-social-science-one

Wardle, C., & Derakhshan, H. (2017). *INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making*. Retrieved from Council of Europe website: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Wauters, B., Thijssen, P., Van Aelst, P., & Pilet, J. B. (2018). Centralized personalization at the expense of decentralized personalization. The decline of preferential voting in Belgium (2003–2014). *Party politics*, *24*(5), 511-523.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: Profile Books.