Let's replay the political debate

De Liddo, Anna; Souto, Nieves Pedreira; Plüss, Brian

Published in:
International Journal of Human-Computer Studies

DOI:
10.1016/j.ijhcs.2020.102537

Publication date:
2021

Licence:
CC BY-NC-ND

Document Version
Peer reviewed version

Let's Replay the Political Debate: Hypervideo Technology for Visual Sensemaking of Televised Election Debates

Anna De Liddo, Nieves Pedreira Souto, Brian Plüss

Knowledge Media Institute, The Open University, Milton Keynes, UK
University of A Coruña, Spain
Centre for Argument Technology, University of Dundee, UK

Abstract

Despite the widespread proliferation of social media in policy and politics, televised election debates are still a prominent form of large-scale public engagement between politicians and the electorate during election campaigns. Advanced visual interfaces can improve these important spaces of democratic engagement. In this paper, we present a user study in which a new hypervideo technology was compared with a publicly available interface for television replay. The results show that hypervideo navigation, coupled with interactive visualisations, improved sensemaking of televised political debates and promoted people's attitude to challenging personal assumptions. This finding suggests that hypervideo interfaces can play a substantial role in supporting citizens in the complex sensemaking process of informing their political choices during an election campaign, and can be used as instruments to promote critical thinking and political opinion shifting.

Keywords: Type your keywords here, separated by semicolons ; Sensemaking, Public Deliberation, Political Election Debates, Hypervideo, Advanced Visual Interfaces, Interactive Visualisations, Deliberation Within

1. Introduction

Considering the widespread public attention devoted to social media and their role in political events, one may wonder whether televised election debates are still as influential today as they were in the past. The 2017 Hansard Society's Audit of Political Engagement in the UK[1] shows that television is still the most prominent means of disseminating political information to a wide variety of citizens – especially to those who lack access to web technologies or are excluded from social media debates due to a lack of interest or digital literacy (Graham & Wright, 2013). Televised elections debates are still important opportunities for democratic engagement, but at the same time are increasingly hard to make sense of for citizens. They present information

[1] The Hansard Society. Audit of Political Engagement Audit 14. The 2017 Report. Last accessed in October 2018 at: https://www.hansardsociety.org.uk/projects/research/audit-of-political-engagement

of uncertain reliability, and are often seen as passive watching experiences, that remain disconnected from the many channels of available political participations - such as social media and media commenting spaces.

There is an imbalance between the potential of televised elections debates to be valuable opportunities for democratic engagement, and people's need to reflect and make sense of political complexity in a more holistic way. For instance, by scrutinizing political dynamics, evaluating facts and evidence, and understanding politicians' claims, also in relation to social media information and the wider political deliberation context.

A gap must be bridged then between internal reflection and public discussion spaces, by the development of individual socio-technical spaces for sensemaking, that can at the same time improve the understanding of the televised political debate and inform people participation in the social media debate.

The research presented in this paper aims to address this gap, by presenting and evaluating Democratic Replay, a new technology which aims to provide a space for citizens' individual reflection, critical thinking and improved understanding of the televised debate. One of our underpinning working assumptions is that improved individual sensemaking can enable better informed political choices and indirectly enhance the quality of public deliberation during an election campaign.

1.1. Proposed Solution and Testing hypothesis

Democratic Replay is an interface for augmented video replays. It uses hypervideo to generate viewing experiences augmented by interactive visualisations. Hypervideo is a video stream that contains embedded anchors, which can be either inside or outside the video's visible area, and are hyperlinked to different types of complementary content: e.g. text, audio, infographics, dynamic visualisations, other video streams, etc. (Smith et al., 2000). Democratic Replay is a hypervideo platform in which these anchors in the video hyperlink to dynamic interactive visualisations, resulting in new interactive hypervideo viewing experiences.

The rationale behind Democratic Replay is that political understanding and reflection of the televised debates can be improved by providing augmented replays, that people can watch at their own pace, and by following their personal interests. Augmented replays add a layer of visual analysis of the debate's content, which shows hidden dynamics, emerging patterns, and aggregated results.

We hypothesize that augmented replays enable users to go beyond what can be perceived watching a simple video, and effectively improve their capability to reflect and make sense of televised election debates.

Our higher-level research questions are: Can augmented replays enhance sensemaking of televised election debates? And to what extent does improving sensemaking positively affect the quality of democratic deliberation?

We address the first research question by using Democratic Replay as an example technology for augmented replay. We present experimental research to test our hypothesis that Democratic Replay enables overall better sensemaking of televised election debates compared to publicly available interfaces for television replay. Finally, we discuss the potential implications of our findings on future research and provide insights on the second research question.

In the remainder of the paper, we describe the experimental study carried out with a panel of 113 people in order to assess the impact of Democratic Replay on personal reflection on and sensemaking of political debates. Results of the study evidence that the new hypervideo and visual analytics interface provided by Democratic Replay significantly improved sensemaking of televised election debates and changed the way in which people shaped their political attitudes and choices during the election campaign.

Even though the main usage of Democratic Replay tested in this paper is the individual exploration of televised debates, our findings have implications on the design of collaborative systems. The study enhances

our understanding of the role that hypervideo-like technologies can play in improving democratic deliberation, and suggests new insights for the design of future technologies for public deliberation that take into account the need of a space for internal reflection and sensemaking.

2. Related Work

2.1. Interactive Systems for Public Deliberation: Motivating the Need of Technologies for Internal Reflection and Sensemaking

Political deliberation is, by its very nature, a collaborative practice that involves individual reflection and public engagement in social debate. Contemporary deliberative democracy has given prominence to the discursive component of this practice, which focuses on citizen participation within political decision-making processes (Ackerman, 1989; Anderson, 2011; Habermas, 1984).

Following the deliberative turn in democratic theory, many CHI and CSCW systems for public deliberation have investigated new technologies for: (1) online deliberation (e.g. the system by Kriplean et al. (Kriplean, Morgan, Freelon, Borning, & Bennett, 2012), its improved version with support for factchecking (Kriplean, Bonnar, Borning, Kinney, & Gill, 2014), and the political deliberation tool by Semaan et al. (Semaan, Faucett, Robertson, Maruyama, & Douglas, 2015)); (2) opinion sharing (e.g. OpinionSpace (Faridani, Bitton, Ryokai, & Goldberg, 2010)); and (3) collective argumentation (e.g. Deliberatorium Iandoli et al 2018), and Cohere (Iandoli, Quinto, De Liddo, & Buckingham Shum, 2014)). These technologies have proven useful in supporting groups of citizens to collaboratively improve the quality of public discourse by, for instance, increasing the quality of learning (Towne, Technology, 2012, n.d.), enhancing political information browsing (Robertson, Wania, Abraham, & Park, n.d.), and favouring constructive rather than confrontational discourse (Kriplean et al., 2014).

Nonetheless, research indicates that deliberation starts in the minds of people who participate in public debate. This is referred to as 'deliberation within':

> "Internal-reflective processes are also involved in responding to the arguments and evidence presented by others in discussion. Much of the work in understanding what others are saying, whether in a formal meeting or an everyday conversation, inevitably occurs inside our own heads." (Goodin & Niemeyer, 2016)

'Deliberation within' is a form of internal deliberation that consists of three main components: understanding one another, internal reflection and imagining discourse (Goodin, 2003). The main idea behind it is that, since in mass deliberation not all participants can take part in open discussion, people embark in a variety of processes of internal deliberation in which they: (1) try to understand other people claims by putting their self in their shoes, (2) they reflect on what are the implications of other people thinking on their understanding of the world, and (3) they shape an opinion. This final opinion shaping is thought of as a fictional statement in the public discussion: basically, people imagine what they would say if they were part of the conversation. In a nutshell, deliberation within is the internal reflection process of imagining multi-party discourse, which happens in people's heads but is social in nature, as it leads to shaping our opinions before we present them in public. This internal reflection process can support the realisation of changes in people's attitudes and opinions before they (or when they cannot) take part in a public debate.

An empirical study of jury deliberation showed that attitude and opinion changes are due mostly to prior internal deliberation rather than to moments of public discussion (Goodin & Niemeyer, 2016). Follow on research has also proved that 'deliberation within' is positively correlated to political interest, political efficacy, systematic information processing, issue interest, issue comprehension, need for cognition, and need for evaluation (Weinmann, 2017). This suggests that, in addition to external, collective processes of discursive

interaction for the realisation of authentic deliberative democracy (Landemore & Mercier, 2010), internal deliberation provides useful spaces for people to think systematically and critically about the issues at stake, and promotes opinion shifting and more active engagement in political activities.

Hence, when designing systems for political deliberation, in addition to spaces for discursive expression and public debate, democratic deliberation technologies must also support spaces for internal reflection, in which people can listen, think and make sense of contents, even before they engage in political discussion. Sensemaking technologies have a longstanding tradition in CSCW but they have not been previously applied to political deliberation. Therefore, in the next section we review the sensemaking literature in CSCW, contextualise it to our application field, with the purpose to define our approach to sensemaking of televised political debates, and motivate the design principles we followed to develop DEMOCRATIC REPLAY.

2.2. Sensemaking: Relevance, Limitations, Proposed Approach and Design Principles

Sensemaking is the process, both internalised and externalised, individual and collective, that people go through to frame, organize, connect, and structure information to understand a problem or situation that they are experiencing. The sensemaking tradition in CHI and CSCW research is longstanding (Bansler & Havn, 2006; Grasso & Convertino, 2012; Russell, Pirolli, Furnas, Card, & Stefik, 2009), and, for many years, researchers have investigated the ways in which technologies can be designed to support the way people make sense of large amounts of data or complex tasks. Several types of behaviours and systems have been explored in CHI, CSCW, and the organisational literature to understand, model, and investigate sensemaking. Russell et al. (Russell, Stefik, Pirolli, & Card, 1993) defines sensemaking as a cyclic process in which people search for external representations of a complex information context or large body of information, and then create, change, and discard these representations with the purpose of either reducing the time, reducing the costs, or improving the quality of a task. Klein et al.'s (Klein, Moon, & Hoffman, n.d.) data/frame model also relies on the coding of information into a frame or external structure that helps encoding new knowledge, identifying gaps, and advancing sensemaking by changing the frame to better fit the context that is being made sense of.

The main common denominator of all sensemaking models and theories is the presence of a representation: a frame or schema that works as a boundary object of reflection and hypothesis testing. The sensemaker iteratively compares her internal views, ideas, and hypotheses with this external representation of the problem. If the two representations match, she experiences a sensemaking moment; if they mismatch, she experiences a "surprise", a knowledge gap. This motivates the generation of a new hypothesis and, in turn, new internal and external representations that are compared in a new sensemaking cycle (see Figure 1 in (Russell, CHI, 2004, 1AD)). In all sensemaking models, the schema or frame is created by the sensemaker, or by a group of sensemakers (Pirolli, on, 2005, n.d.), who share a common understanding of the problem, and it is built with a process of information seeking, clipping, and annotation.

Annotations represent the crucial step of making unstructured information into some form of meaning or knowledge (L. Nelson et al., 2009). Many sensemaking technologies have therefore consisted of tools supporting information markup and annotation. Such tools also support the structuring and visualisation of annotations. For instance, NoteCards (Halasz, Moran, & Trigg, 1986), one of the first sensemaking technologies, Spartagus (Hong, Chi, Budiu, Pirolli, & Nelson, 2008) and ClaimSpotter (Sereno, Shum, & Motta, 2005), are tools used to annotate and make sense of web documents. Some sensemaking tools have also used a mixed-initiative approach in which human and machine annotations are combined to support sensemaking (De Liddo et al., 2012).

Sensemaking is a very relevant process to many collaborative work practices, still sensemaking technologies have failed to demonstrate substantial impacts on collaborative knowledge works, and their uptake has never scaled. One of the main challenges to user adoption is formalisation (Shipman and Marshall, 1999). Sensemaking is a process of constructing meaning that involves the placing of information into a schema or frame (Weick, 1995). This formalisation process comes with a cognitive overload, which is often not proportional to the anticipated benefits. Additionally, most sensemaking technologies target expert analysts and rely on manual annotation of large data corpus, thus putting a high workload on the sensemaker, and often requiring specific expertise and training. In order to make full use of the advantages of having an external representation to support making sense of complexity, users need: to learn a new formalism first, then follow it to structure, clip, label and express their thinking. All these extra steps to be taken before to act are barriers to usage (Shipman and Marshall, 1999).

Experiences with a variety of formalisms (from hypermedia, to argumentation and design rationale, knowledge–based systems and general groupware) show that users find formal interactions difficult. Formalisation brings about a series of barriers to cooperative work, such as: premature data structuring, management of users' disagreements, and slowing down of generation tasks to allow introspection. The lack of widespread adoption of sensemaking tools has also been associated with the complexity of the user interfaces for building and structuring annotations, as well as the difficulty for non-experts to make sense of the data model behind the sensemaking schema.

In the attempt to overcome these shortcomings, Russell et al. (Russell, Jeffries, CHI, 2008, n.d.) argue that sensemaking technologies need to be simpler, embedded in normal practice, and use more intuitive visualisations. Shipman and Marshall (1999), in their comprehensive analysis of the limits of formal representations in interactive systems, suggest five design criteria for system designers to minimize problems of formalisations:

- only formalise what is essential to the core supported task;
- make sure that the costs of formalisations are well matched by the benefit it brings;
- enable incremental formalisation and restructuring;
- provide ephemeral structure on demand; and
- support users with training and facilitation.

Following these recommendations, we adopted three main principles in the design of a technology for sensemaking of televised political debates replay:

1. *DP1 - aid to formalisation*: we leverage external aid from experts (and/or automated machine analysis) for the phase of clipping, indexing and annotation of the raw data from the televised debate, to build the external representation (the augmented replay) at no costs to the final users
2. *DP2 - structure on demand*: various formalisations, and their associated representations (the interactive visualisations), are made available to the users on demand and without additional cognitive effort.
3. *DP3 – simple formalisation embedded in normal practice*: the interactive visualisations are embedded in the normal flow of the task (in our case the replay of a video), and are simple and not disruptive.

Following these design principles, in our approach to sensemaking of televised political debates, the annotation and schema generation are performed by expert analysts or machine analyses, leaving only the sensemaking loop of hypothesis generation, testing, and communication to the non-expert sensemaker. Intuitive visualisations are then built to improve the sensemaking process for non-expert users. In this sensemaking approach, visualisations play the same role as Klein's frame (2006) or Russell's representation (Russell et al.,

1993), and provide structured narratives to support reflection and understanding; they work as 'inferred structures' which cost the final users very little compared to what they provide.

In the study we report in the following, we focus on a set of expert analysis and annotations of televised election debate videos (DP1), from which we designed interactive visualisations that can be explored on demand to augment the video replay (DP2). The overall goal is to improve sensemaking by providing 'frames' of interpretation and understanding of the televised election debate.

Given the nature of our sensemaking problem, we looked at hypervideo as the most promising technology to build and communicate simple interactive visualisations that evolve together with the video replay (DP3). In fact, hypervideo both enables the aggregation of multiple analytic data in a single video visualisation (DP2); and it is not disruptive of the viewing activity, since multiple infographic layers can be rendered at runtime (DP3). Hypervideo is a new media that has been only partially studied and applied to the context of political communication. Therefore, in the next section, we position our research in the hypermedia, television interfaces, and political communication literature, and we highlight how our study contributes new knowledge to the understanding of the affordances of new hypervideo technologies to make sense of televised political debates.

2.3. New Television Experiences for Political Sensemaking: Interactive Hypervideo Visualisations

In this era of ubiquitous Internet use and handheld computing devices, television remains the most popular source of political information. According to Ofcom (The Communication Market Report, 2019), people spend 22.4 hours per week watching television, as well as considerable unrecorded time discussing and acting upon what they see. And, according to the Hansard Society's Audit of Political Engagement (n. 15) in 2018 69% of the UK population regarded television or radio as their main source of election-related news and information, compared with lower percentages for the printed press (39%), news websites (32%) and social media (21%). The civic importance of television has long been recognised also at a policy level. The main free-to-air broadcasters in the UK retain public service obligations to produce media content that caters for the interests of citizens not just consumers (Ofcom 2019). Despite its long-lasting relevance, and the perception often associated to it as an "old" media, television is evolving, both technologically and in the ways viewers experience it. The HbbTV 2.0 industry standard (Van Deventer et al., 2013) will enable television sets and handheld devices to synchronise over contents, leading to enhanced, interactive viewing experiences (e.g., via companion mobile apps) and creating potential for novel applications (Bibiloni, Mascaro, Palmer, & Oliver, 2015; César Garcia & Geerts, 2016).

Technological innovations bring new opportunities for widening participation outside usual television audiences, for instance, by attracting new viewers via social media (Anstead & Ben O'Loughlin, 2011; Enslin, Pendlebury, & Tjiattas, 2001). New industry standards for TV broadcasting also provide further interactive and immersive viewing experiences (Anstead & Ben O'Loughlin, 2011) which can attract new audiences. However, these new viewing experiences can also introduce disadvantages, as relatively complex digital skills are required to interact appropriately with televised contents (Coleman, 2012; McLeod & Perse, 2016). The challenge is then to find a balance between effortless usage and explicative power of new interactive experiences with televised programs to improve engagement and sensemaking in political debates. We attempt to reach this balance by augmenting the experience of watching televised political debates with interactive hypervideo visualisations, that is to say interactive visualisations combined with hypervideo replay.

Recent advancements in the use of hypervideo, applied to the analysis of users' behaviours in web navigation, suggest that interactive hypervideo visualisations can improve understanding, by incorporating the temporal information in the rendering of the visual analytics (Leiva and Vivó, 2012). This result shows promise

in the use of interactive visualisations as tools to improve hypervideo interactions. Still, Hypervideo and interactive visualisations have scarcely been used in combination.

On one hand, hypervideo has been evolving after it was first introduced in 1996 with the design of HyperCafe (Sawhney, Balcom, & Smith, 1996). Since then, the hypervideo tradition in the hypertext research community has grown with innovative applications such as Advene, a tool for supporting active reading with hypervideo (Aubert & Prié, 2005). Systems for deploying hypervideo on the web nowadays, such as Popcorn.js and the HTML5 video tag allow deploying sophisticated video annotations via any web browser. Most research is focused on teleconferencing or learning applications (Chambel, Zahn, Finke, 2004, n.d.; Sawhney et al., 1996); and most hypervideo applications consist of manual annotation and hyperlinking of either text or static visualisations overlaid on the video. Interactive visualisation overlays to hypervideo are less common.

On the other hand, several attempts have been made to use interactive visualisations to improve sensemaking in the context of political debates. Threaded conversation interfaces have been used to represent the debate's arguments and dynamics (Snell, 2010). Argument maps (Renton & Macintosh, 2007), and deliberation dashboards have been used to show content and social network dynamics (Iandoli et al., 2014). A recent proof of concept with the BBC Moral Maze program has focused on argument visualisation of a radio debate on voluntary abortion[2]. This application of interactive visualisation to BBC radio demonstrates broadcasters' interest in real-life experimentations with visual analytics to improve the content and navigation of complex debates[3]. Still, no applications of interactive visualisation overlays to televised programme has been carried out so far. Additionally, research indicates that interactive visualisations do not always deliver on the promise of improving users' understanding of the complex information which they distil (Keim et al., 2008). Users with lower digital skills struggle to interact and most importantly make sense of visual analytics outcomes (Pantazos, IVAPP, 2012, n.d.). This points to the fact that Interactive Visualisations are difficult to be made sense of, even when considered as stand-alone interactions. Coupling this interaction with hypervideo replay can potentially increase barriers, rather than enhancing sensemaking. This calls upon more empirical research on the affordances and impact of interactive visualisations to improve sensemaking, both, in general, and more specifically when coupled with hypervideo.

Despite the advancement of television experiences and hypervideo research, there is a lack of hypervideo technologies designed for television interaction and a lack of empirical studies which compare these new visual interfaces with existing mainstream television replay interfaces to understand the affordances of these technologies. Also, there is a lack of understanding of the effect of hypervideo interactions on political debates, specifically, the capability of hypervideo replays to improve sensemaking, enhance critical thinking, and affect political attitudes and choices.

This further motivates the testing of our hypothesis that: Democratic Replay hypervideo technology enables overall better sensemaking of televised election debates compared to publicly available interfaces for television replay. To test this hypothesis, we set up an experiment to systematically compare Democratic Replay with a common replay interface for televised election debates. To allow formal comparison between the two interfaces, we used a 9-factor framework, that we drew from the literature, to formally measure sensemaking improvements.

---

[2] Centre for Argument Technology. Piloting Argument Technologies with the BBC. Last accessed in October 2018 at: http://bbc.arg.tech/bts/.

[3] Visit http://bbc.arg.tech/ to explore the interactive analytics developed for the voluntary abortion debate.

2.4. Measuring sensemaking

Measuring sensemaking is full of subtleties (Russell et al., 1AD). In much of the sensemaking literature, sensemaking is measured with either quantitative approaches looking at things such as task completion time, task accomplishment, or number of documents retrieved, or complex multiple-choice tests to assess the rapid understanding of large document collections (Russell et al., 1AD). Completion times and task accomplishment have proved to be inaccurate indicators of sensemaking (O'Day & Jeffries, 1993). For instance, people read at different speeds, and bring different domain knowledge to the sensemaking task which makes them act faster (Spink, Wilson, Ford, Foster, & Ellis, 2002). Therefore, using time as a tool to compare sensemaking behaviours may bias the results.

Alternatively, researchers have also taken qualitative approaches based on the assessment of textual summaries. In these latter cases, sensemaking is measured on the basis of the quantity and quality of acquired knowledge, by extracting clues on the level and complexity of insights from the narrative's syntactical and semantic analysis (M. J. Wilson & Wilson, 2013). These methods focus on experts' judgement and require time-consuming discourse analysis. They prove very useful to study sensemaking behaviours and processes, but are less useful to systematically compare different sensemaking outcomes.

Alsufiani et al. (Alsufiani, Attfield, & Zhang, 2017) suggested a more generalizable quantitative method for measuring sensemaking, which is based on the construction of survey questions measuring five theoretical features of sensemaking: 1) gaining insight; 2) reducing confusion, uncertainty, and ambiguity; 3) finding connections; 4) structuring; 5) gap-finding and gap-bridging (Table 1).

Table 1. 9-factor Sensemaking framework: Sensemaking features, their alternative definition, and theoretical provenance.

N.	Sensemaking Factors from the Literature	Adapted Definition of Sensemaking factors - for Political Debate
1	Weick: Retrospect	*Reflection*
2	Alsufiani et al.: Gaining insight	*Insights*
3	Alsufiani et al.: Finding connections	*Focus*
4	Alsufiani et al.: Structuring	*Argumentation*
5	Alsufiani et al.: Reducing confusion, uncertainty, and ambiguity	*Explanation*
6		*Assess Facts and Evidence*
7		*Distinguishing*
8	Alsufiani et al.: Gap-finding and gap-bridging	*Assess Assumptions*
9		*Change Assumptions*

We use Alsufiani et al.'s five theoretical factors of sensemaking but propose some changes to contextualise the metrics to our sensemaking process. Two of Alsufiani et al.'s factors (gaining insights, and finding connections by focusing on a diversity of aspects of a problem) can be easily translated to the context of making

sense of a political debate, and have been only renamed as Insight and Focus (factors N.2 and N.3). On the other hand, the three features of: 1. structuring; 2. gap finding and gap bridging; and 3. reducing confusion, uncertainty, and ambiguity; were reviewed in light of their application to our specific sensemaking problem.

Structuring (factor N.4), is the capability of framing a problem. In our case, this can be interpreted as the process of identification, connections making, and mapping of the ideas and arguments raised during the political debate. Chinn and Anderson (Chinn, Record, 1998, n.d.) define this process as argumentation: an interactive process of searching for reasons and evidence for different positions in the form of a conversation. We therefore propose to rename, structuring as argumentation, since this is a better construct to measure structuring behaviours during sensemaking of a political debate.

In the same way, if we think about watching a political debate, Alsufiani et al.'s theoretical factor of reducing confusion, uncertainty, and ambiguity, can be better broken down into three distinctive sensemaking behaviours: reducing confusion by explaining the issues discussed (Explanation factor, N.5); reducing uncertainty by evaluating facts and evidences (Assess Facts and Evidence factor, N6.); and reducing ambiguity by discerning between options (Distinguish factor, N.7 in Table 1).

In the context of making sense of a political debate, Alsufiani et al.'s theoretical factor of gap-finding and gap-bridging, can be translated in the two sensemaking factors of: assessing and changing previous political assumptions (factors N.8-9). These are two crucial sensemaking capabilities we aim to measure. In fact, previous research shows that political election debates are often perceived as leaders winning or losing contexts, in which people tend to polarize their positions. This polarization is a common problem identified both in social media debate and the sensemaking literature and it is traced back to a problem of confirmation bias. Confirmation bias can be defined as people's common attitude not to focus in their analysis on the disconfirmation of their hypothesis (Coleman, Blumler, Moss, & Homer, 2015). Therefore, we consider people's attitude to assess and change assumptions (N.8-9) as key factors to assess the quality of sensemaking in political deliberation.

Finally, we add reflection to our sensemaking framework (factor N.1), as a key factor to measure sensemaking in political debates, and which was not captured by Alsufiani et al.'s theory. Building on Weick's definition of sensemaking, reflection can be interpreted as the capability to retrospect and look back at the implications of what has been said in a political debate, in light of one's own personal understanding and assumptions. For Weick, 'the basic idea of sensemaking is that reality is an ongoing accomplishment that emerges from efforts to create order and make retrospective sense of what occurs' (Weick, 1993, p. 635), therefore thinking back and reflect are key sensemaking capabilities.

On that account, we propose the following sensemaking framework which consists of 9 factors:

1. **Reflection**: as the capability to think back and in depth;
2. **Insights**: as the capability to get unexpected ideas or make unexpected inferences;
3. **Focus**: as the capability to see different angles and aspects in the debates;
4. **Argumentation**: as the capability to reconstruct the arguments that the speakers make;
5. **Explanation**: as the capability to identify and explain issues;
6. **Assess Facts and Evidence**: as the capability to assess presented facts and evidence;
7. **Distinguishing**: as the capability to make a difference between the speakers' claims and the options proposed
8. **Assess Assumptions**: as the capability to assess personal ideas, opinions and assumptions;
9. **Change Assumptions**: as the capability to change one's own mind as a consequence of watching the debate.

We used this 9-factor framework to measure sensemaking improvements in the evaluation of Democratic Replay. In the following we describe the system's main architecture and user experience, before moving the focus to its empirical evaluation.

3. Election debate visualisation: DEMOCRATIC REPLAY

Democratic Replay can be described as a multi-channel hypervideo platform for sensemaking of political debates. It is an online video replay tool, built to offer citizens a new way to watch the UK's 2015 General Election ITV Leaders' Debate. Democratic Replay aims to slow down the political debate, and to reduce sensemaking complexity while watching it: (1) by aggregating multiple parallel information streams (such as fact-checking websites, social media feeds) in one place and under a unified user interface; (2) by providing advanced visual analytics (such as speakers' performance, argument structure, and audience reactions); and (3) by enabling a new interactive and dynamic exploration of the debate video.

Figure 1 shows the conceptual overview leading to Democratic Replay. Data sources related to the election debate (such as video, transcripts, audience feedback, and tweets) are time-linked together with analytics and

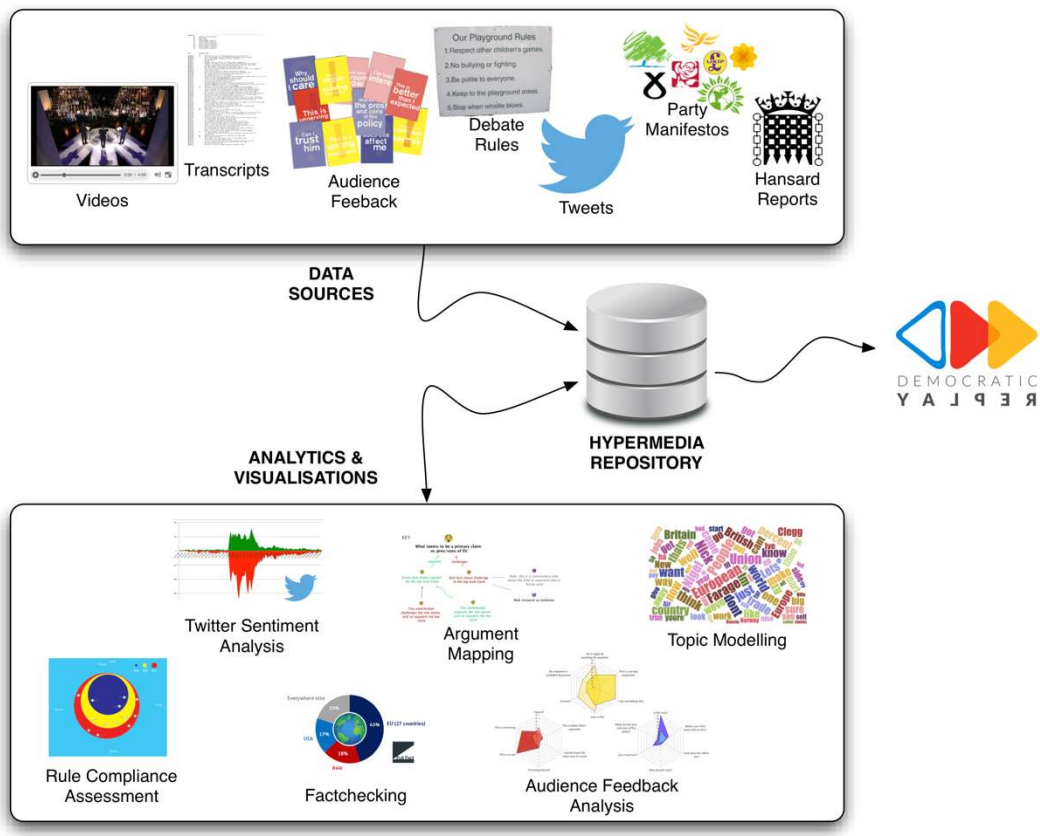

Figure 1. DEMOCRATIC REPLAY is the front end of a hypermedia repository combining heterogeneous data sources, complex analytics, and interactive visualisations.

visualisations (such as argument mapping, factchecking, and twitter sentiment analysis) into a hypermedia repository. These data and analytics are then presented to users seamlessly, providing an interactive viewing experience structured around analytical storylines.

The analytics and visualisations are based on analyses of the data from the sources, performed either automatically or manually by expert analysts (DP1). The argument mapping, for instance, was carried out by a computational linguistics and dialogue mapping expert. The speakers' performance analysis was based on a typology of qualitative judgements on the actions of the politicians, that emerged from the manual data coding of the debate's transcripts carried out by a political science expert. Twitter activity on the other hand, was tracked automatically from the data collected during the debate. Factchecking was also collected publicly, from manual analyses produced by Full Fact4, a UK-based independent factchecker NGO. Audience feedback was collected by a novel method for tracking nuanced audience responses to live events (De Liddo et al. 2017) and analysed automatically.

By carrying out the production of the analytics and visualisations in this manner, we split the sensemaking process in two very distinct types of activities: 1. Clipping, analysis and annotation are done by experts or automatically, while 2. hypothesis generation, testing, evaluation, and communication, are left for the sensemaker (see section 2.2 for design principles and rationale).

3.1. User Experience in Democratic Replay 2.0

We carried out a series of user studies (see for instance Plüss and De Liddo (2018)) and interface redesign cycles to achieve the user experience (UX) of Democratic Replay described and evaluated in the remainder of the paper. A team of seven people, including developers, computer scientists, graphic designers, UX experts, and political scientists, worked intensively for three years to achieve a sensemaking technology that is sophisticated yet intuitive. The objective of this paper is to assess the extent to which we found a balance between the ease with which Democratic Replay can be used, and the enhanced explicative power it provides for making sense of televised election debates.

The design of the UX started with sketching and prototyping of the interactive visualisations based on the main informational elements in each analysis. Initially, we envisaged the user interface as a collection of interactive visualisation panels activated via widgets (see Figure 2).

A first set of visualisations was tested with users via mixed workshop and focus group sessions (Pluss and De Liddo, 2018). The results of this evaluation indicated that, although those interested in politics saw the attractiveness, appeal, and potential of Democratic Replay 1.0, they found the visualisations too complex to be of use without proper introduction to their visual language and overall meaning. As a result of this evaluation we re-targeted our end users from all citizens to citizens with at least some interest in politics, and we redesigned the visualisations to make them more intuitive and self-explanatory. In order to reduce the amount of information presented at once in each visualisation, we structured the data unveiling based on the hierarchical nature of the analytical categories. In the current version of Democratic Replay (2.0) users are therefore able to select the aspects of each analysis on which they want to focus, and to access the visualisations on demand (DP2).

[4] https://fullfact.org/

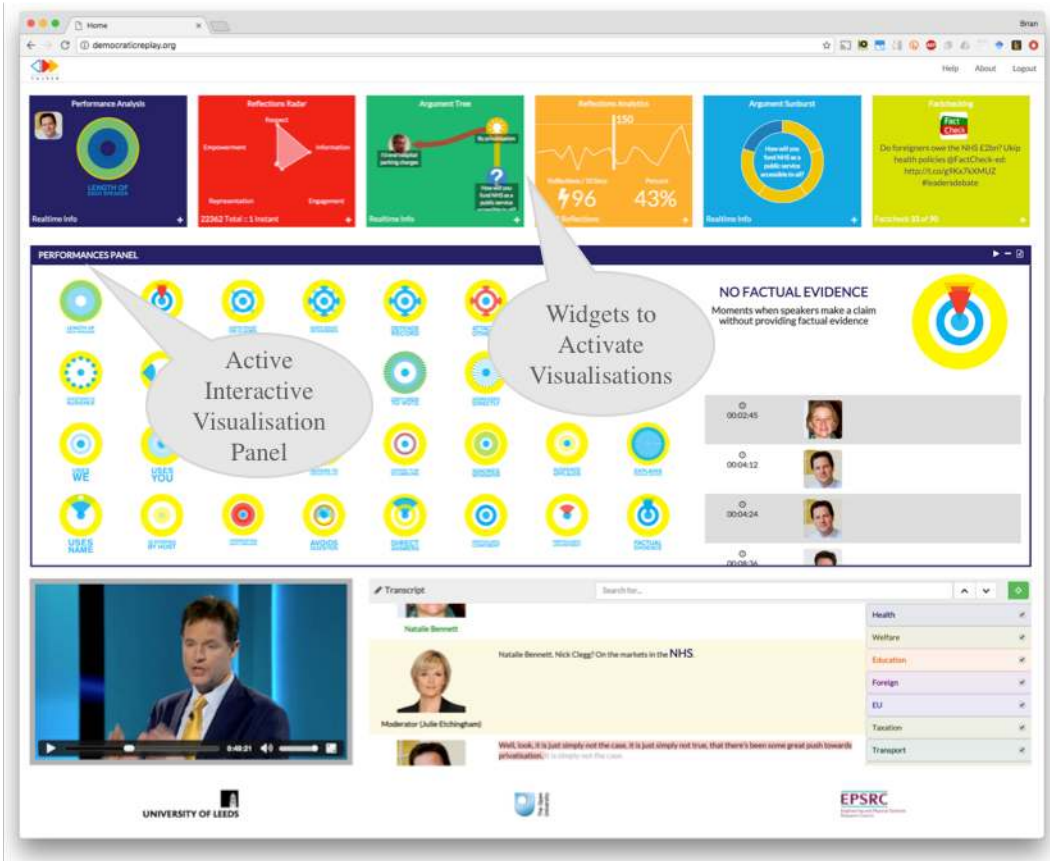

Figure 2. DEMOCRATIC REPLAY 1.0: a collection of widgets (top row) which gave access to further details in the visualisations panel (middle row).

The same principle was applied to the way visualisations in the system are first introduced to users. The homepage of Democratic Replay 2.0 is structured around a set of intuitive questions directly related to each analysis (see Figure 3; rationale and further details in Section 3.2). Then, a set of sub-questions leads to specific configurations of a visualisation page aimed at providing an answer to the sub-question. Users can then continue to explore other aspects of that visualisation, or go back to the homepage and move on to another type of analysis. The resulting design provides a UX offering the full potential of the system in stages and based around intuitive questions.

Users arrive at Democratic Replay 2.0 by landing on the homepage shown in Figure 3. The top of the page includes a carousel with six slides that briefly introduce the purpose and motivation of the system[5]. This is followed by a set of 7 high-level questions (see coloured squared buttons at the bottom of Figure 3).

[5] The slides cycle automatically in a carrousel and read: (1) Political talk can seem fast. Too fast. What

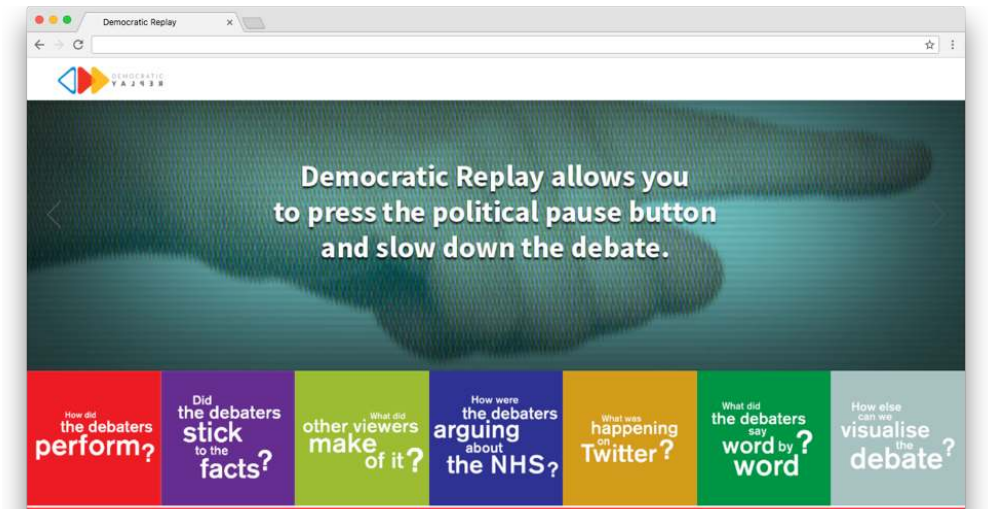

Figure 3. Homepage of DEMOCRATIC REPLAY 2.0 providing a hierarchical user experience structured around questions related to each of the analyses behind the interactive visualisations.

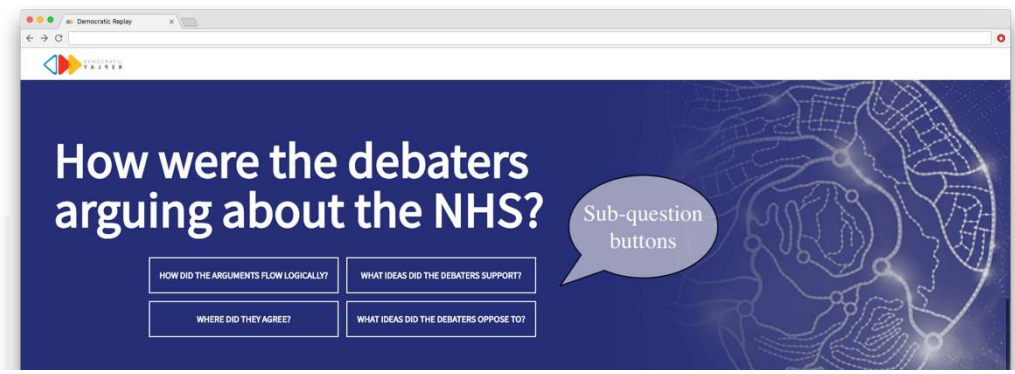

Figure 4. Question Section for the Argument Map. The four sub-question buttons read: How did the arguments flow logically? What ideas did the debaters support? Where did they agree? What ideas did the debaters oppose to?

These questions structure the home page and provide the main entry points for users to navigate the system. We designed the questions by thinking at what most users might want to know in connection with an election

would it look like if we slowed it down? (2) Is this a good argument? What's the evidence? Should we believe what they say? (3) Democratic Replay allows you to press the political pause button and slow down the debate. (4) To make sense of what our would-be leaders are saying. To participate actively and meaningfully in contemporary democracy. (5) During the 2015 UK General Election Campaign we analysed the ITV Leaders' Debate. (6) To help you tackle the complexity of the debate in your own time and allow you to ask the questions that matter to you.

debate. Each question can be clicked and points at a different analysis (and its associated interactive visualisation) in the system. The questions and the analyses they refer to are shown in Table 2. A demo video of the system in action can be accessed online[6] and provides a better sense of the user experience and interaction with the tool. By clicking on these coloured squared buttons or scrolling down the page, users access Question Sections. These sections contain sub-questions that further refine the high-level question. Figure 4 shows the Question Section related to the Argument Map analytic. Clicking on a sub-question button leads to a Question Page, which contains the interactive visualisation, set up in a so that the analysis provides an answer to the sub-question.

Table 2. High-level questions on the homepage of DEMOCRATIC REPLAY, and the analytic to which they relate.

Homepage Questions – User's Navigation Entry Points	**Analytics**
How did the debaters perform?	Performance Analysis
Did the debaters stick to the facts?	Fact checking
What did other viewers make of it?	Audience Reflections
How were the debaters arguing about the NHS?	Argument Map
What was happening on Twitter?	Twitter Activity
What did the debaters say word by word?	Transcript Augmentation
How else can we visualise the debate?	Other Visual Analytics

Figure 5 shows the Visualisation Page for the Argument Map, after a user clicks on the 'How did the Arguments Flow Logically?' sub question.

After clicking, the sub-question turns white (white indicates selection), and a panel with the interactive elements answering the selected question opens below the sub-questions panel. This shows the entire available argument map for the discussion around funding the NHS (Figure 5). Users can then explore the debate further in two main ways: from the analytics to the video, or from the video to the analytics.

They can use the interactive elements in the visualisation to jump to different moments in the debates. When an interactive visualisation element is clicked, the video jumps to the moment in which the element clicked becomes relevant and starts playing. At this point the transcripts of the debate synchronously scroll down with the replay, and show in red highlight the instant statement (bottom-right text box in Figure 5). Alternatively, users can explore the visualisation through the video replay. They can simply click play on the video box (top right in Figure 5). At that point the video starts, the transcripts scroll down, and the interactive visualisation dynamically shows and hides information that is relevant to what is being said in the moment. Users can then return to the homepage, by clicking on the system logo or using the back button, which restarts the UX.

[6] A 3 minutes demo video of Democratic Replay is available at: https://youtu.be/lc9htNNNQdo

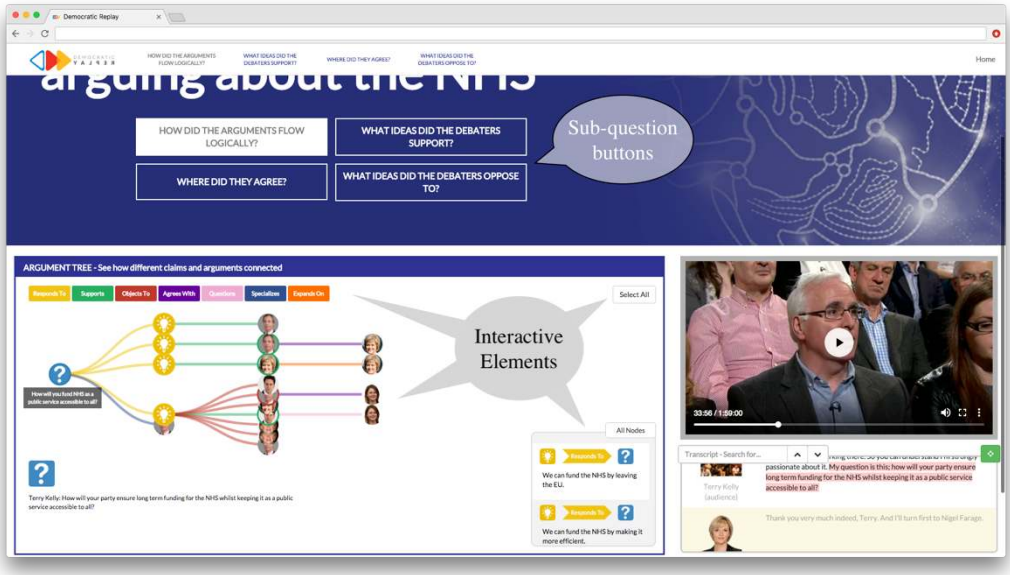

Figure 5. Question Page for the Argument Map showing all available logical connections. The header replicates the Question Section in the homepage (see Figure 4). Coloured buttons, speaker bubbles and list entries in the visualisation panel (bottom left) are interactive elements which allow users to navigate the video non-linearly.

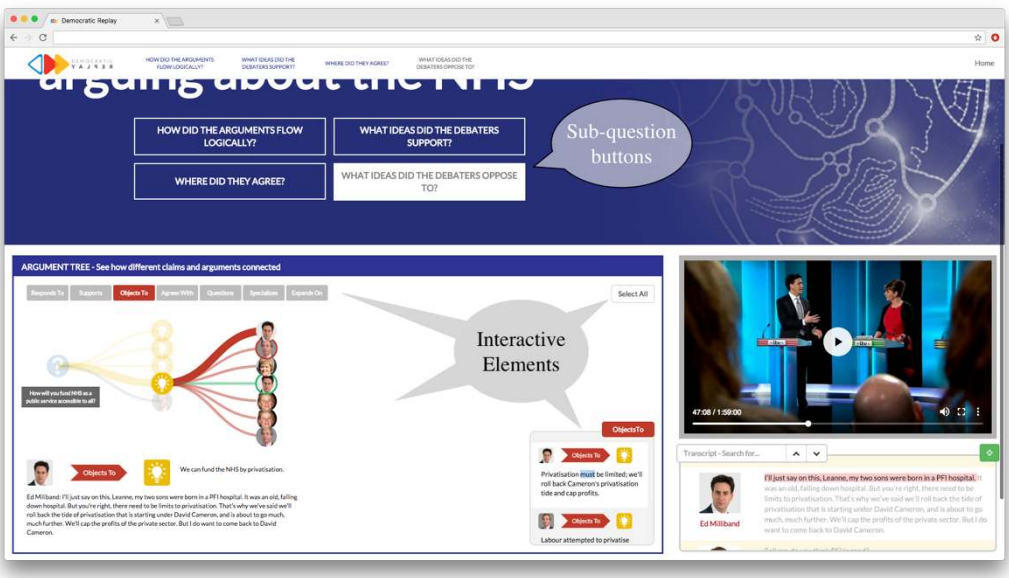

Figure 6. Question page for the Argument Map, showing the ideas debaters opposed.

Figure 6 shows the Argument Map visualisation set up to answer the sub-question 'What ideas did the debaters oppose to?'. Sub questions are used to provide filtered views. For instance, in Figure 6 the interactive visualisation shows only the contrasting ideas and arguments in the map (red links in figure 6), and greying out the rest of the information.

To give another example, let's look at another question accessible from the Homepage (Figure 3): 'How did the debater perform?'. This takes to the Question Section for the Performance Analysis (Figure 7). Here too, by clicking on the sub-question buttons, users can access specific configurations of the interactive visualisation.

Figure 7 shows the Visualisation Page for the sub-question 'Did they provide evidence?'. In this new interactive visualisation textual round buttons are used to indicate moments of relevance – for instance 'Factual Evidence Missing', or 'Casting Doubt on Others' Claims' in Figure 7 – classified by the qualitative typology of actions of the politicians emerging form the expert political discourse analysis. Only the moments' types that are relevant to the selected sub-question (in white in Figure 7) are shown at once. As users replay the debate, performances indicated in the analysis at the time of playback are highlighted with a grey square pop-up hint, which then shades away after 2 seconds (see 'Factual Evidence Missing' circle in Figure 7). Users can also explore occurrences via the list under Moments, and jump to that point in the video by double-clicking on an entry. The list can also be filtered by debate participant, using the pictures at the bottom of the panel.

The mechanism for accessing and navigating the other analytics and visualisations is analogous to the above and can be also explored by watching the video demonstration7.

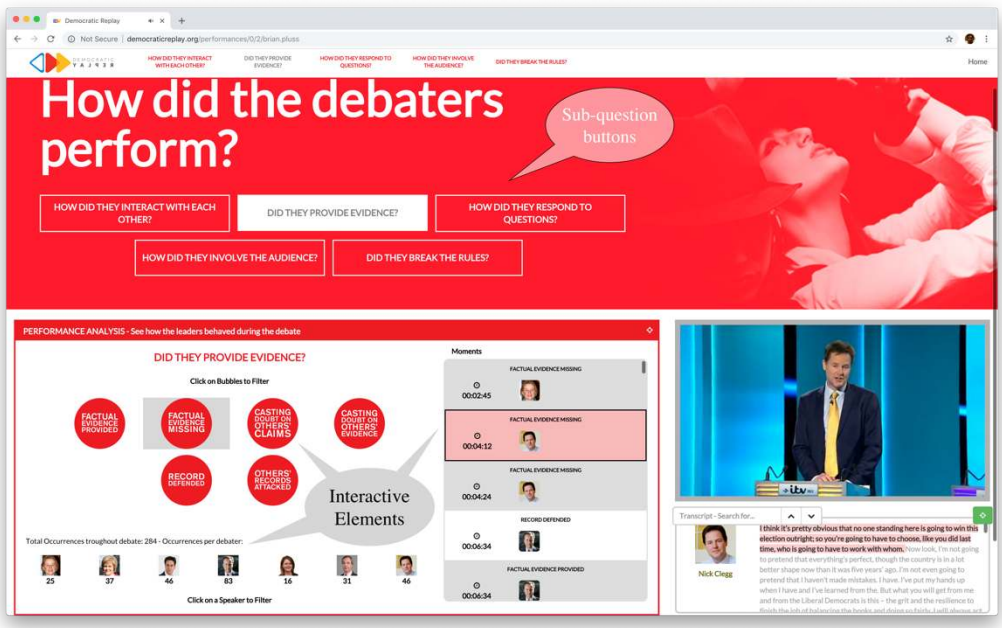

Figure 7. Visualisation Page for the Performance Analysis showing the badges related to the provision of evidence by the politicians during the debate. Users can navigate the visualisation by using the sub-question buttons in the header, or by clicking on the interactive elements in the visualisation, which allow users to filter moments and navigate the video non-linearly.

[7] A 3 minutes demo video of Democratic Replay is available at: https://youtu.be/lc9htNNNQdo

4. Empirical Evaluation During the 2015 UK General Election Campaign

We tested Democratic Replay in the context of the Prime Ministerial debates of the 2015 UK General Election. Specifically, we used a two-hour event called the ITV Leaders' Debate, involving the leaders of the UK's 7 main political parties, and aired live on April 2, 2015.

We chose the ITV Leaders' Debate for its political and civic relevance, and its uniqueness in terms of bringing the main political figures of the election campaign face-to-face for extensive debate. The anticipated citizens' interest in this event, made it the ideal opportunity on which to focus our technology testing. In fact, it was later on reported that over seven million people watched the debate live[8]. Subsequent research confirmed that the debate reached a broad audience, with a surprising 48% of survey respondents who identified as 'not very interested in politics' having watched the debates (Coleman et al., 2015). The crisis of engagement, both in terms of voter turnout and of how citizens make sense and assess the options available to them, meant that enhancing the single most popular political event of a General Election campaign was a timely effort. During the debate, we collected the media file, carried out the pre-processing, live data aggregation, post hoc analysis, transcriptions and expert annotations, necessary to develop the interactive visualisations. We then set up Democratic Replay ready for testing.

4.1. Research Method and Experimentation Design

The aim of our experimental research study is to show how to what extent the user experience provided by Democratic Replay has a direct causal inference on users' reported sensemaking capabilities. As stated in the introduction, our research hypothesis is that Democratic Replay enables overall better sensemaking of televised election debates compared to publicly available interfaces for television replay. To test our hypothesis, we carried out an A/B testing between two independent groups of users.

A/B testing is an approach commonly used in controlled online experiments to generate design insights (Deng et al. 2017, Kohavi and al. 2007). It is a method widely adopted both by large corporations (Kohavi et al. 2009), as well as HCI theorists (Hancock et al. 2007) and by applied researchers to demonstrate the utility of a technology (MacKenzie and Zhang 1999). It is often used to assess if one UI is better than another in usability studies, and in general for comparing user experiences and manipulations of cognitive or technical variables of interest. Experimental research practices like A/B testing "can be applied to make HCI more rigorous, informative and innovative" (Gergle and Tan 2014, pp.194). As a methodology, experimental research has a variety of advantages over other HCI methods: it produces quantitative data that can be analysed using inferential statistics, it can be replicated and extended to other contexts to progress research knowledge, and it allows to minimise researchers' biases and other systematic errors. Of course, experimental research has also its limitations, such as the risk of low external validity, as the capability of research claims to hold true in other contexts or settings (e.g. ecological validity). In our study, we followed recommendations by Olson et al. (1993) in order to maximise external validity of our study by: choosing a realistic task (that is watching an online televised debate during an election campaign); testing in a natural setting; and selecting participants that are

[8] The Guardian (2015). 7m people watched leaders' debate – ITV. Last accessed in October 2018 at https://www.theguardian.com/politics/2015/apr/03/7m-people-watched-leaders-debate-itv-seven-etchingham

representative of the general public during political elections (with a variety of age, sex, work, and demographics).

We followed a between-subject design and compared the overall user experience of two independent groups of users: one group used the BBC News replay interface (control) to watch BBC Leaders' debate, and the other group used the Democratic Replay interface (treatment). We chose the BBC News replay interface (Figure 8) for baseline comparison, as it was the only election debate–specific and publicly available replay platform during the election campaign; and because, being the BBC a mainstream broadcaster in the UK, we expected it to be the most common replay interface used by UK citizens.

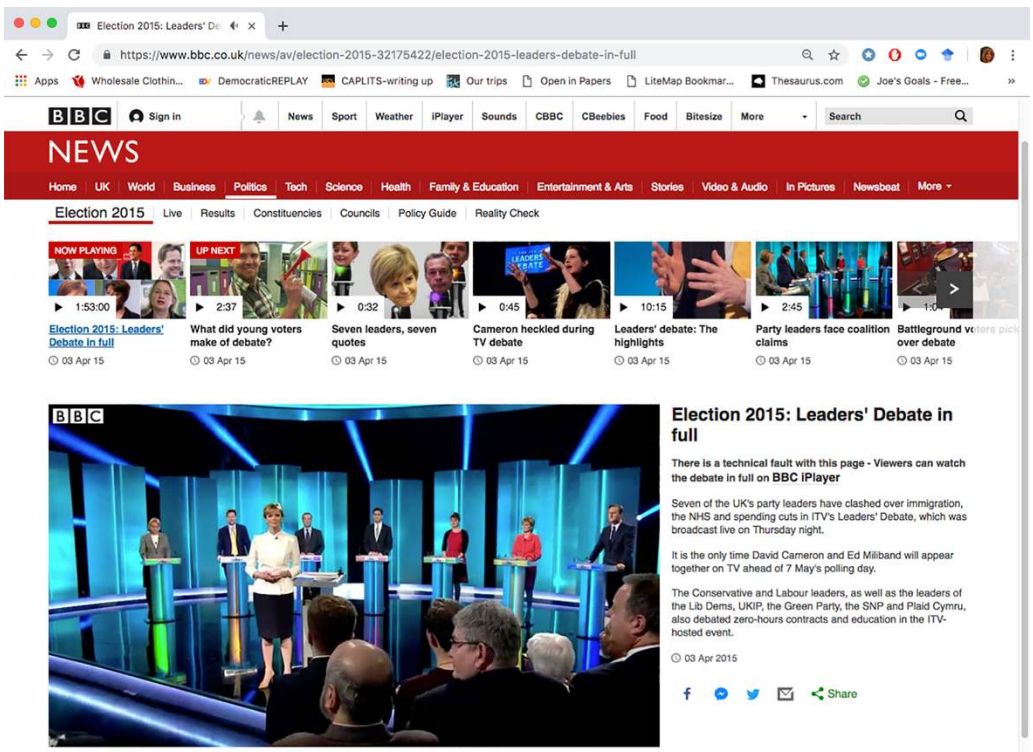

Figure 8. BBC News replay interface of the 2015 ITV Leaders' Debate.

The two user interfaces and experiences we compared (the full version of Democratic Replay, Fig 3-9, and a mainstream BBC replay interface Fig.10) are quite different. We choose to compare these specific systems for two main reasons. Firstly, by comparing a new unfamiliar (rather complex and potentially overwhelming) tool with a tool of everyday usage, if we found positive results in terms of improved sensemaking with Democratic Replay, we'd be more likely to measure effects independent from the training, usability or familiarity confounding variables. Secondly, as mentioned above, to improve external validity we carried out the experiment in a natural setting, with a tool of common usage, during election time, with a relevant televised election debate, and in the usual setting in which replay technologies are naturally used (that is to say, in an online replay setting).

4.1.1 Participants

To improve the power of our statistical conclusions, we drew our insights from a large sample of participants. We involved a total of 113 people, who took part either face-to-face or through an online website. This was mainly an economic choice: face-to-face recruitment via a market research recruiting agency was rather costly, to stay within project funding limitations, and in order to double the number of participants, we recruited and engaged additional participants via Mechanical Turk.

All care was taken to ensure that the different recruitment modalities did not affect the internal validity of our experiment. Specifically, user interaction with the experimental setting was exactly the same for both groups, with no additional knowledge or expertise required or provided based on modality of participation.

The face-to-face sessions took place in a computer lab at the University of Leeds, UK, in March 2017. Recruitment was through a local market research agency (r3search: http://r3searchleeds.co.uk). Due to lab capacity, the sessions took place over two days (20 and 21 March) with two sessions of 14/15 people on each day: one with the control group and one with the treatment group. We requested 60 people in total (2 in the treatment group eventually did not attend). As conditions for recruitment, we requested a gender and age (18-65) spread across the groups. The incentives for participation were £25 per hour and all participants were from the Leeds metropolitan area in the North of England.

Same characteristics of gender and age group spread, and same incentives, were applied to online participants recruited on Mechanical Turk, the only difference being that online participants were recruited outside the Leeds metropolitan area, with a good national spread. We had participants from: Aberdeenshire, Berkshire, Cambridgeshire, Devon, Essex, Fife, Gloucestershire, Greater London, Hampshire, Hertfordshire, Inverness-shire, Kent, Lanarkshire, Lancashire, Leicestershire, London, Northamptonshire, Nottinghamshire, Oxfordshire, Staffordshire, Surrey, Sussex, Yorkshire and other.

Participants in the control group used the BBC News replay interface to replay the political debate, and the treatment group, used the Democratic Replay interface. In order to ensure that control and treatment groups were not unequal because of the different recruitment modalities we used the same ratio of online/face-to-face recruited people in both treatment and control groups. In the control group of 57 people, we had a 30/27 face-to-face/online recruitment split. While in the treatment group of 56 people, we had a 28/28 face-to-face/online recruitment split.

In order to allow research conclusions that are independent from key demographics, treatment and control groups were also balanced in terms of age and gender with a nearly equal split per demographic category between the two groups. Overall, a total of 52 participants were women (46%), 60 men (53%), and 1 preferred not to say. The age range was 18 to 65, with subsets of 45 people (40%) between 18 and 30, 32 (28%) between 31 and 40, and 36 (32%) aged 41 or more.

Crucially, the only conditions for selection were that (a) *participants knew how to use the Internet*, and (b) they answered affirmatively the question: '*If there were to be a televised leaders' debate at the time of the next general election, would you choose to watch it*?'. As our focus was on evaluating technology, we requested that participants expressed an interest in televised election debates but did not consider political inclination. This selection's question was only added to filter out people with no interest in politics. In fact, empirical evidence from previous research {Pluss:2018jz} showed that people uninterested in politics, who would not have watched an election debate in the first place, would not engage voluntarily with an election debate replay experience, with or without technological mediation. Voluntary participation is a key requirement for authentic civic engagement, therefore, in the recruitment of our panel we excluded people that would not realistically want to use the technology even if it were available to them. We should therefore assume that our findings can be only generalised to people with at least some interest in politics and that would voluntarily engage with televised

elections debates in the future. We also acknowledge our recruitment being skewed toward people living in the Leeds's metropolitan area, nonetheless, as the focus of this paper is not on drawing political conclusions, but on evaluating fundamental cognitive capabilities enabled by technology mediation, there are no reasons to believe that a skewed geographical distribution of participants may be a confounding variable in the study.

We paid particular attention to the screening process for engaging Mechanical Turk subjects being identical to the procedure used by people engaging face-to-face. The compensation for participants' work was also the same. All subjects received 25 dollars per hour, a rather larger compensation compared to average Mechanical Turk work. We hoped that the larger compensation, in addition to the description of the task as highly cognitive demanding, and the focus on the impact of the work in advancing research, would motivate Mechanical Turk participants to perform quality work. Additionally, to make sure that only faithful worker's responses were kept, we added a content question in the middle of the process, this is a question that required writing a textual summary of the content watched, and could be appropriately replied only if the participant really engaged with the viewing experience. Also, various open questions were added at the end of the process. The richness and details of the answers to these questions was considered when evaluating the trustworthiness of the overall engagement with the experiment. All these questions were manually assessed by the research team to carry out the appropriate data sanitisation.

4.1.2 Experimental Setup

The experiment was set up in three phases: 1) a pre-interaction phase, in which, before watching the debate, users were asked to fill in a questionnaire designed to gather information on general demographics, digital literacy, interest in politics, and pre-existing knowledge and assumptions that might affect the sensemaking and engagement measures; 2) an interaction phase, to allow time for participants to explore the platform; and 3) a post-interaction phase, in which participants were asked to fill in a post-interaction questionnaire to measure effects on participants' reactions in terms of sensemaking of the political debate. An online wizard was built to guide participants through the three main experimentation phases, the same wizard was used by face-to-face and Mechanical Turk participants.

In the interaction phase, all participants were introduced to the assigned interface (BBC News replay interface or Democratic Replay) by a five-minute free exploration of the site to gain expertise with the interface. After that they carried out a task-oriented exploration of the website. The task consisted of finding out about the claims and positions of the Leader of the Green Party during the debate, and then write a textual summary about what they found. Participants were given 10 minutes to freely explore the website to solve this task. The textual summaries were used only to evidence engagement with the cognitive task, and the content was not analysed to assess tasks' result. The Green Party was chosen because it is a minority party in the UK, and therefore more likely to resonate neutrally in people's attitude toward the task.

In the post-interaction phase, data were collected using a digital questionnaire. The post-exploration questionnaire was designed to quantitatively assess sensemaking effects of the user experience. The questionnaire included a series of five-point Likert scale questions, which informed the statistical analysis. In the following we describe the analysis we carried out, and discuss the results of this analysis.

4.2. Data Analysis

We carried out a quantitative data analysis of the survey data and used the nine-factors sensemaking framework proposed in section 2.4 to build a series of five-point Likert scale questions to measure people's self-declared capability to make sense of the political debates (a list of the questions can be found in Appendix

1). Some of the questions were framed in a positive style, and some of them in a negative one, this was done to avoid a pattern of responses, and to lead respondents to maintain closer attention to the questions. Before the analysis, every answer to the negative questions was inverted to be compared with the other questions in the same polarity scale.

4.2.1 Scale Reliability

To measure internal consistency of the sensemaking scale, as represented by the 9 proposed sensemaking features, we calculated McDonald's Omega. McDonalds Omega (McDonald, 2013) has been shown by many researchers to be a more sensible index of internal consistency – both in relation to the well-established Cronbach's alpha (Cronbach, n.d.) and when compared to other alternatives. Omega has less risk of overestimation or underestimation of reliability and does not require unidimensionality between metrics. Since in our case unidimensionality is unsure and multidimensionality is suspected (especially considering that most psychometrics measures, like our sense making factors, are likely to be related in some form), Omega is a more appropriate test to assess internal consistency of the sensemaking scale (Dunn, Baguley, & Brunsden, 2014). Omega allows to check the extent to which correlation between factors can undermine scale reliability by calculating the impact on Omega Confidence Intervals (CIs) of each of the metric individually. Following (Dunn et al., 2014) procedure, we have calculated McDonald's Omega along with CIs for the aggregate 9 items scale, and repeated the calculation by omitting one item at a time. We have then identified the subset showing the highest Omega value with the narrowest CIs, which represents the subset with best reliability. To do the computation we used JASP, a free statistical software from the University of Amsterdam (*JASP (Version 0.13))*. We used the desktop user interface, distributed under GNU Affero GPL v3. This software permits to establish the confidence interval (CI) and to select the method for the estimation. One way to find omega value is do a factor analysis of the original data set, rotate the factors obliquely, factor that correlation matrix, do a Schmid-Leiman (schmid) transformation to find general factor loadings, and then find omega. An acceptable value of reliability for omega coefficient is between .70 and .90 (Campo-Arias & Oviedo, n.d.), although in some circumstances values higher than .65 can be accepted.

4.2.1 Hypothesis Testing

Since our analysed data is nominal and normal distribution cannot be assumed, we chose a non-parametric test. The statistical analysis was carried out with a Mann-Whitney U test, a modification of the Wilcoxon ranking method (or Wilcoxon signed-rank test), to compare the sensemaking effect between the two groups. As we had large samples (more than 30 subjects), when applying the Mann-Whitney U test, the value of U approaches a normal distribution (Bellera:2017gt}), and so the null hypothesis was tested by a Z-test.

We calculated effect sizes with Cohen's interpretation for Pearson's product-moment correlation r (Cohen, 1988), which is appropriate to assess effect sizes in non-parametric tests. For classifying positive effects, we followed Cohen's suggested average values for: small effects (r=0.10), intermediate effects (r=0.30), and large effects (r=0.50). With ranges from 0.25 to 0.35 being the most common correlations reported in the literature between discriminable different psychological variables (Cohen, 1988)). We therefore used the following conservative ranges for small effects (less than 0.25); medium effects (0,25 to 0,35); and large effects (more than 0.35).

4.3 Results and Discussion

Results of the scale reliability analysis show excellent Omega values. With a confidence interval of 95% the point estimation for omega is 0.908 (being the CI lower bound 0.883 and the upper bound 0.933) (See Table 3). Results also show that the aggregate scale is better than any other combination, which means that all the items add relevant information to the experiment. If we look at Omega posterior mean value for the aggregated scale, we can see that this is higher than any combination obtained by dropping each item at a time (Table 4). We also calculated the CIs difference and confirmed that the aggregate scale has narrower CIs than the second best (with Debate Assessment item dropped, last raw in Table 4).
The full set of 9 metrics shows the highest Omega with the narrowest CIs, hence confirming reliability of the 9 factors sensemaking scale beyond threats to unidimensionality.

Table 3. Results of the Omega and CIs for aggregated 9 factors scale.

Scale Reliability Statistics	
Estimate	**McDonald's ω**
Posterior mean	**0.908**
95% CI lower bound	0.883
95% CI upper bound	0.933

Table 4: Results of the Omega and CIs for the scale if each item is dropped.

Individual Item Reliability Statistics			
	McDonald's (if item dropped)		
Item	**Posterior Mean**	**95% CI Lower bound**	**95% CI Upper bound**
Focus	0.895	0.866	0.922
Assumptions	0.899	0.871	0.925
Reflection	0.897	0.868	0.924
Insight	0.894	0.865	0.921
Explanation	0.892	0.861	0.92
Distinguish	0.897	0.87	0.925
Argumentation	0.901	0.872	0.926
Personal Assessment	0.894	0.866	0.922
Debate Assessment	0.904	0.877	0.93

Results of hypothesis testing show that for 7 of the 9 factors were improved by the hypervideo technology. Table 5. Results of the Mann-Whitney U test for the nine sensemaking factors. Rows in grey indicate the sensemaking behaviors significantly improved by Democratic Reply. reports in detail the values of Mann-Whitney U value, p value, and Z-value for every one of the nine sensemaking features. For seven of them, we rejected the null hypothesis (see grey rows in Table 5. Results of the Mann-Whitney U test for the nine sensemaking factors. Rows in grey indicate the sensemaking behaviors significantly improved by Democratic Reply.). We measured the product-moment correlation r to assess the direction and the sizes of the reported effects. Every obtained value of r was found positive, which means that there are no adverse effects. We can therefore conclude that 7 out of 9 of the measured sensemaking behaviours were significantly improved by the use of Democratic Replay.

Table 5. Results of the Mann-Whitney U test for the nine sensemaking factors. Rows in grey indicate the sensemaking behaviors significantly improved by Democratic Reply.

Sensemaking Feature	**U**	**p**	**z**	**r**	**Effect Size**
Reflection	1175.00	0.0107	2.5534	0.2402	Small Effect
Insight	**1113.50**	**0.0031**	**2.9590**	**0.2784**	**Medium Effects**
Focus	1270.50	0.0422	2.0317	0.1911	Small Effect
Argumentation	1244.50	0.0321	2.1431	0.2016	Small Effect
Explanation	1325.00	0.0917	1.6862	0.1586	/
Evaluate Facts & Evidence	**1032.50**	**0.0006**	**3.4203**	**0.3218**	**Medium Effects**
Distinguish	1385.50	0.2050	1.2675	0.1192	/
Assess Assumptions	1257.00	0.0409	2.0444	0.1923	Small Effect
Change Assumptions	1203.00	0.0177	2.3711	0.2231	Small Effect

The highest effects were found for the capability to Evaluate Facts and Evidences, with effects in the higher end of the medium effects zone (Table 3. Results of the Mann-Whitney U test for the nine sensemaking features). This sensemaking capability is linked to Alsufiani et al.'s factor of reducing uncertainty. In fact, reducing uncertainty can be seen as a stimulus for sensemaking that occurs when there is a lack of knowledge (Alsufiani et al., 2017). People experience uncertainty when they do not know what information to trust, and therefore seek facts and evidence to reduce this uncertainty. Sensemaking is our way to build 'contextual rationality' in those uncertain situations, by building from 'vague questions, muddy answers, and negotiated agreements that attempt to reduce confusion' (Weick, 1995); Weick, 1988). Our findings demonstrate that Democratic Replay provides an instrument to face uncertainty problems, by providing significantly better 'ways to evaluate the facts and evidence' presented during the political debate (QS6 in Appendix).

This finding directly links to the growing attention to misinformation and fake news in deliberative democracy and media research (Guess et al. 2018, Fourney et al. 2017). Citizens feel too often manipulated and increasingly unable to distinguish facts versus speculations in a public debate; and seek new ways to cope with uncertainty. Our finding indicates that representing the political debate in a way that better organize, structure and visualize its dynamics, provides sufficient support for citizens to increase confidence in their ability to discern between facts and speculations. We can argue then, that in situations in which uncertainty cannot be avoided, or facts do not exist or are too hard to be confidently assessed, sensemaking technologies are a viable

option to enhance citizens' confidence in their ability to evaluate facts and evidence and to make better informed political choices.

Democratic Replay also outperformed the BBC News replay interface in the capability to generate new Insights, with medium effects (Table 3). A significantly higher number of people in the treatment group declared that they gathered 'unexpected insights on the debaters and what they said', thus contributing an element of surprise to the viewing experience. Surprise is a key sensemaking component (Weick, 1995), and can be seen as a strategic political element, which can trigger people's interest and curiosity to seek new political information. This result is particularly promising if we think about the potential impact of new hypervideo technologies as political engagement tools. In fact, unexpected insights and 'serendipity' are increasingly recognised as key capabilities to develop in technology mediated communication to go beyond existing political debate shortcomings such as lack of diversity, groupthink and polarisation (Sunstein, 2018).

Democratic Replay was also found to significantly improve Reflection and Changing Assumptions (at the higher end of small effects, Table 3). A significantly higher number of people in the treatment group declared that they could 'reflect on the debate in a deeper way' and most crucially, 'changed some initial assumptions they had before the debate' (QS1 and QS9 in Appendix).

This is a key finding for us, since it demonstrates the capability of sensemaking technologies to support people to go beyond win-lose thinking and help them to better shape, and even change their opinions. Enabling people to reflect and change assumptions was one of our main motivators for designing spaces for internal deliberations. The evaluation study confirmed our hypothesis that designing for reflection and sensemaking can promote critical thinking and opinion shifting, thus counteracting negative dynamics in political engagement, such as confirmation biases and polarisation (Golbeck et al., 2017; Taber & Lodge, 2006). There is a link here with Kahneman's (Kahneman, 2012) fast and slow thinking modes in contemporary politics. As argued extensively (e.g. by (STOKER, HAY, & BARR, 2016)) fast, intuitive and visceral, thinking is ubiquitous in many democracies. Mainstream media, pundits and candidates engage citizens in fast-paced, constantly moving messages with little time for reflection. Slow thinking is required for deliberation and "processing" political discourse. Technology that helps the transition and sustainment of slow thinking improves the democratic quality of the experience. So Democratic Replay is intended as a way to slow down election debates, giving citizens/viewers the time and tools to think slowly, rationally and deliberatively. Our results indicate that the system is in the right direction. However, a larger, *slower*, evaluation study one-to-one semi-structured interviews and qualitative analysis is needed to assess this more robustly.

To a lesser extent (small effects), but still significantly better than the BBC News replay interface, Democratic Replay helped citizens to assess their personal assumptions and to 'reconstruct the arguments that the speakers made' (QS8 and QS4 in Appendix). This is encouraging, in that it confirms an overall improvement in people's argumentation and personal assessment capabilities. It is surprising though that, despite argumentation analysis being the object of one specific interactive visualisation (Figure 4-6), the impact on capability to reconstruct arguments was not in the wider effects zone.

The smaller positive effect was recorded for the capability to 'focus on different aspects of the debate'. We speculate, this being due to the nature of dynamic interactive visualisations, which constantly change the focus of attention, distracting the viewers from the flow of the replay. Simplicity and minimisation of disruptions in the viewing experience was one of our design principles (DP3), to which we owe to pay closer attention in future redesign.

Finally, the two sensemaking features that did not provide significant results relate to the tool's capability to reduce confusion and ambiguity (Distinguish and Explanation features). The mean difference of the scores shows that there is a slight, but not significant, improvement of these two features in the treatment group.

The quantitative analysis cannot shade light on the reasons for the different performance of the 9 sensemaking features, a qualitative data analysis would be needed to better understand and triangulate our quantitative findings.

5. Implications for Future CSCW Research Directions

Citizen disengagement from politics is one of the main issues in modern democracy. It can be argued that highly unexpected election results, such as Brexit in the UK, have been affected by a lack of political engagement (only 51% of the population voted for the Brexit referendum, and young people in a lower percentage). This shows the importance of adding new ways to involve people in policy and politics and also improve the way in which people can make sense of the complexity of political situations during an election campaign, including, and most importantly, ways of understanding the effects that specific political decisions may have on their life. Sensemaking technologies can be used to tap into new internal motivations for people to take part in political debate, they can appeal to their desire to understand and decide for themselves. To allow this process of sensemaking, we need better tools for people to seek, extract, and interpret relevant information, and we need usable, user friendly, and intuitive devices to make sense of the political complexity, during the political events in which these policies are presented to the public and critically discussed (such as during televised election debates).

This study has shown that Democratic Replay improved users' capability to make sense of the political election debate when compared to a common TV replay interface, by enhancing 7 out of 9 sensemaking factors. Obviously, one of the limitations of our study is that the sensemaking scale we used may not cover all the spectrum of sense making capabilities, as there is no general consensus at present time on what the prefect set of capabilities would be. However, we have shown independence and high internal consistency reliability of the nine factors. While they might not cover all the aspects, they represent distinctive aspects of sense making, and 7 out of those 9 have been significantly improved by the hypervideo visualisations technology. Specifically, the study shows that this new type of engagement with televised elections debates improved citizens' sensemaking capabilities such as: 1) reflecting, 2) focusing on different aspects of the political debate, 3) gaining new insights, 4) reconstructing the arguments that politicians are making, 5) assessing facts and evidence, 6) evaluating personal assumptions and 7) changing them as a result of the viewing experience. These capabilities are fundamental for deliberative democracy, and are part of the key democratic functions of televised election debates as expressed by citizens (Coleman & Moss, 2015). Citizens expect televised election debates to go beyond mere channels for the delivery of political information, and seek them to enable important democratic functions, such as 'to evaluate political claims and make informed decisions', or 'to make a real difference in the political world' (Coleman & Moss, 2015).

On that account, we distilled three main principles that we believe can be crucial to the realisation of deliberative democracy in today's public communication landscape.

1. Design for Deliberation 'Within'. If we want to support people's capability to question assumptions and think critically, we need to design spaces for personal reflection and sensemaking. Political deliberation does not start nor finish in the public sphere; on the contrary, it is intertwined with our deeper thinking, goals, values, and beliefs. Platforms for public discursive debate need to be coupled with platforms for internal reflection, assessment, and sensemaking of the debate. Previous research indicates that social presence also means fear to express opinion (Rainie, Project, 2012, n.d.) and designing for citizens' engagement also requires designing for comfortable 'social distance' (Lampinen et al., 2017). In the same way, designing for public deliberation also means designing for private spaces of intimacy and introspection, otherwise called 'deliberation within.' Gordon and Manosevich also refer to this as 'in-person' deliberation (Gordon & Manosevitch, 2010), and

suggest that it can support qualities such as trust and rationality, which are usually difficult to establish in common public deliberation contexts. The hypervideo platform presented in the paper is an interface for 'deliberation within.' It is a system to support the internal reflection and sensemaking process involved in political deliberation, which takes place in people's minds while they take part in political events, or before they take part in public discussions through explicit verbal engagement. Our findings confirm recent research evidence that these spaces of internal deliberation effectively and significantly enhance opinion shaping and shifting (Goodin & Niemeyer, 2016).

2. Design for Collective (Human-Machine) Sensemaking. Individual sensemaking processes need human–machine support. One of the main obstacles to the adoption of sensemaking technologies is the complexity of the tasks, data annotations, and visualisations that need to be generated for the sensemaking process to be realised. Additionally, visualisation aids are usually designed for expert analysts rather than lay users (Russell et al., n.d.). Our research shows that in the context of making sense of televised debates some of the sensemaking steps, such as data annotation and representation tasks, can be effectively delegated to machines, or other forms of human experts' analysis. The non-expert sensemaker can still improve personal sensemaking and understanding, even if he does not personally go through data annotation and structuring. This finding establishes an important precedent on the exploration of new mixed-initiative approaches to computer-supported collective sensemaking of public debates. Recent advances in research on argument mining

(Lawrence & Mining, 2015; Lawrence & Reed, 2020; Lippi & Torroni, 2015) and automatic detection of fake news (Shu, Sliva, Wang, Tang, & Liu, 2017) are starting to shape a near future in front of us in which machine analysis will be able to assist human assessment of complex political debates and discourse dynamics. Those technologies are excellent candidates to automate some of the key human expert analysis that we carried out in our study. Automation also means immediacy. This implies that the augmented replay of the televised election debates would be available during or immediately after the event, which would exponentially multiply the impact of the technology on opinion shaping before the political vote. This touches on the sensitive and very timely impact of technology in politics and democracy. Susskind's notion of "Future Politics" (Susskind, 2018) places technology, "powerful digital systems", at the core of the way politics and democracy will happen in the near future. As ubiquitous technologies and "the algorithm" become increasingly more involved with the directions our (private and public) lives take, such systems become political. Their designs, contents and implementations, and effectively the policies they enforce, must be discussed, negotiated and regulated. The engineering behind such systems is no longer software engineers, but also social engineering. Although much smaller in pervasiveness than social network, financial market, and personal banking systems, Democratic Replay is social software in that sense and must conform to scrutiny and "earn" the trust of the users. That is, the legitimacy of the knowledge within and the righteousness with which the system guides the users in their sensemaking must be accountable. As one of the participants in our study informally expressed after the session, *"academics drive the analyses that make for the system's data, but why should we trust academics?*", This is a valid point, and our solution to this, in addition to aiming for increasingly more systematic and automated analyses away from human expert biases, is to offer the system under the triple standard of open platform, open source, open data. This allows for Democratic Replay to be scrutinised (both in data and code), extended with ad hoc channels (such as we did with factchecking data in the version described here), and redeployed. 3. Design for Data Aggregation and Visual Analytics. One of the main features provided by Democratic Replay is the capability to aggregate multiple sources of data and different types of analysis of the debate, which are coherently integrated and represented in usable, intuitive, and interactive visualisations. This feature was designed to address two main issues identified in the literature: dispersion of the debate (Semaan, Robertson, Douglas, & Maruyama, 2014), platform islands and community isolation (Klein, & Convertino, 2015). Dahlgren points out that political deliberation forms a 'sprawling public sphere' (Dahlgren, 2009); this means

that people discuss and share relevant information with a variety of social media channels which produce silo conversations. New tools are needed to bridge political debate across community platforms. Within the online deliberation research community, this has been done by creating new online deliberation technologies for public debate. We propose a different approach in which new hypermedia and sensemaking platforms are designed to support data aggregation, structuring, and visualisation to provide citizens with tools to analyse and make sense of existing political debate, also across different media platforms (such as television and online media). In this context, the role of visual analytics interfaces is paramount in reducing information complexity and enhancing sensemaking of large amounts of data and complex political debates.

6. Limitations

In this paper, we devised and tested a tool to improve sensemaking of televised election replays targeted to non-expert users, specifically general citizens with at least some interest in politics. From a theoretical perspective, we have split the sensemaking processes in two: i. a first phase of clipping, indexing and annotation of the raw data to create interactive visualisation, which is performed by expert analysts; ii. and a second phase of hypothesis generation, testing and communication which is carried out by the non-expert sensemaker while replaying the political debate. This choice was made as a strategy to overcome some recognised limits of formal representations in interactive systems, specifically by providing inferred structures to support reflection and understanding (DP1). However, this split in attributing sensemaking tasks may have effects that we have not anticipated and owe to be explored further. What do we lose, in terms of capability to make sense, when the non-expert sensemaker does not perform the annotation himself? What is the trade off in terms of sensemaking improvements? Does the split of sensemaking responsibilities create problems of legitimacy of data analysis? What are the trade-offs between reducing formalisation efforts and producing accountable interpretations?

This study shows that Democratic Replay improved sensemaking of political election debates, and therefore an approach to sensemaking in which different users perform different tasks is viable. Still, we are not able to assess to what extent, and what are the potential negative implications of such an approach. For instance, the role of the human/machine analyst is per se political, since it overlays layer of external interpretation, that should be itself subject to scrutiny, and may become object of manipulation. Future research should look at the socio-political aspects of the technology and its potential use and impact in real life politicised contexts. Issues of trust and accountability of data analysis, when representations are inferred by expert or machine analysis are also worth exploring in the future. It would be interesting to find out to what extent, enabling larger user's control, can improve trust in expert or machine analysis. For instance, new design solutions could be required to enable users to check, validate or challenge expert analysts' interpretations.

Another limitation of the study is that it is focused on assessing the general user experience, which consists of the aggregation of seven interactive visualisations, each one complex in its own way. Therefore, our findings are limited in the way we can associate sensemaking improvements to specific design choices. For instance, we can claim that Democratic Replay improves the capability to change personal assumptions, but we do not know which one of the visual analytics proposed contributed more to this improvement. We know that Democratic Replay improved the capability to assess fact and evidence but we do not know if this improvement is prominently due to the presence of a factchecking aggregator (Factchecking visualisation), or if it was rather due to the capability to assess speakers claims (Performance Analysis), or analyse the debate in a more structured way (Argument Tree). An assessment of the affordances of each of the 7 interactive visualisations (Table 2) would improve the understanding of context and details of our findings.

Finally, the study presented in this paper aimed at drawing conclusions that may be generalised in terms of demographics and target audience. Still, we know that nowadays citizens require 'democracy on demand'

(Coleman et al., 2015). This means that democratic spaces and rights cannot be interpreted uniquely and should respond to people's personal needs, interpretation and understanding of society. Technologies for democratic public deliberation need to be designed with a variety of democratic entitlements (Coleman & Moss, 2015) in mind, and they need to be customized to the needs of different demographic groups if they aim to reach all citizens. Therefore, future studies should focus on demographic analysis, showing to what extent hypervideo technologies like Democratic Replay appeal to different demographic subgroups with different sensemaking behaviours. For instance, it would be interesting to investigate to what extent different people need different public deliberation spaces to make sense of a public debate.

One aspect that we did not address in the studies reported here is that of the current levels of political discontent, public disengagement and generally anti-politics, with the subsequent surge in populist, "apolitical" candidates. At the outset of the Election Debate Visualisation project, Democratic Replay was intended as tool to increase user engagement with politics. However, the first evaluation study we run on an early version of the system (see (Plüss & De Liddo, 2018)) showed that it did not induce an interest in politics for users who were previously disengaged. Generally speaking, study participant appreciated the potential of the system, but said they were simply not interested in politics and would not invest the effort required to engage with the visual analytics. One participant drew a comparison with the football match analytics that nowadays feature in most sport television shows: "if offered to someone who is simply not interested in sport media, they might appreciate the potential and usefulness, but will not necessarily be made more interested in or engaged with sport shows". This realisation was behind the selection of people who already have an interest in politics and would watch a televised election debate for the studies reported in this paper.

What technologies would increase engagement is still an open question and the subject of much needed further research. Mainstream media and current politics are taking the ""fast thinking"", emotionally loaded approach for engagement, but as discussed above, this is not the best way to improve the quality of political discourse, engagement and democracy.

7. Conclusion: Summary and Contribution

The research presented in this paper provide preliminary evidence of the positive role that sensemaking theory and technologies can have in improving informed participation in public deliberation. We have shown that Democratic Reflection, a sensemaking tool which combines interactive visualisation with hypervideo navigation, crucially provide new ways for citizens to make sense of political debates, check facts and evidence, gain new insights, and more confidently assess and change their assumptions.

Our findings suggest that using new media technologies specifically designed to improve critical thinking and sensemaking, like Democratic Replay, can have a considerable impact on informing and shifting the opinion of voters, when applied during a televised election campaign. As initially argued, television is still important for participation in politics, but it is now mixed with a variety of other media. This produces complex information spaces that are overloaded, uncertain, unclear, and therefore alienate citizens from quality engagement with political events. In this paper, we have hypothesized that the issue of citizens disengagement from politics is not only in the platforms used for public debate, but also in the lack of sense making support for individual reflection to meaningfully contribute to the debate.

We argue, that democratic deliberation technologies must also support spaces for internal-reflection, in which people can listen, think and make sense, even before they engage in political discussions. Public deliberation is political, therefore the essential requirement to express opinions in a public sphere is unquestionable. Nonetheless, in order to reflect and change their mind, people need spaces where they feel free from peer-pressure, social judgement and power relationships, spaces where is easier to think slower, admit

being wrong and change. These individual, private and safe space are spaces for deliberation within. Goodin and Niemeyer have proposed panels of experts questioned on prime-time television as key events for internal deliberation (Goodin & Niemeyer, 2016). Similarly, televised election debates can be interpreted as events for people to individually reflect and make sense of the political debate, witness different perspectives, and gather insights and evidence to inform their political opinions and choices. These individual sensemaking and assessment capabilities are key democratic functions of deliberation in the public sphere, and if well mediated, can improve people's understanding of political discussions. To test this hypothesis, we have proposed Democratic Replay, a tool for individual sensemaking of political election debates, which we have compared with an existing media platform for television program replay (the BBC News replay interface). From this comparison, we have learned that sense making technologies can effectively support the understanding of complex political debates, thus potentially making them more accessible to general citizens. Sensemaking tools can trigger people's curiosity, enable the gathering of new insights, and support people to assess and change their assumptions. In line with the growing research interest in technologies, to check the quality and reliability of information (Shao, Ciampaglia, Flammini, & Menczer, 2016), our findings show that sensemaking tools have the potential to support people to think critically and better inform their vote. From this we have drawn three design principles for future technologies to improve public deliberation and sense making of public debate, that are: enable spaces for blended internalised/externalised reflection and debate; use machine power to infer structure rather than explain; and create intuitive visualisations to navigate complexity.

The paper makes four main contributions to research. Firstly, a 9-factor sensemaking framework has been proposed. We have shown sufficient independence and found a high level of internal consistency reliability between the nine sensemaking features introduced at 2.4. This empirical evidence supports the adoption and future exploration of the proposed 9-factors sensemaking framework in political communication research contexts. Of course, the framework needs to be further validated in other contexts and settings. Secondly, we have provided evidence that, when fact checking is unavailable or too hard to be carried out, like in contexts of fast evolving political debates, sensemaking tools can help people tackle problems of uncertainty in data interpretation, and improve their confidence in evaluating facts and evidence. Thirdly we have provided a positive example of how, if provided with the right tools, people can go beyond a win-lose interaction with political debate, can think critically, and show an active capability to re-assess and change positions. In particular, visual analytics narratives and hypervideo navigation have shown to trigger changing of personal assumptions that people hold before watching the debate. This is a very encouraging result, which addresses the ongoing concern about the real value of new media in the context of political debate and democratic deliberation: specifically, the scepticisms towards their capability to support people's critical thinking rather than promote polarization of pre-existing groups and opinions. Finally, this paper contributes to research on new television experiences by showing the successful application of an hypervideo tool to improve sensemaking of televised political debates. Our study can be used to motivate further research on hypervideo technologies, as a tool to effectively augment televised experiences with dynamic visual analytics. In this context, we demonstrated that coupling interactive visualisations with hypervideo provides an analytic narrative overlay which significantly improved sensemaking and exploration of the video experience.

Acknowledgement:

This research was funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant EP/L003112/1. DEMOCRATIC REPLAY has been designed and developed by the authors, and is the result of 4 years of intense research, development and testing. The co-authors contributed to the paper as follows: Anna

De Liddo: 70% (conception, design, interpretation, writing), Nieves Pedreira Suoto: 20% (analysis and wring up of the statistical study), Brian Pluss 10% (acquisition of data, system description writing, revision of manuscript).

Appendix A.

Please find in the table below pre and post exploration survey's questions.

Pre-exploration Survey:

QD1 (Age): Age ([18-24] [25-30] [31-40] [41-50] [>51])
QD2 (Gender): Gender ([Female] [Male] [Prefer not to say])
QD3 (Decided Vote): When the 2015 General Election campaign started, had you decided on how to vote? ([Yes] [No])

Post-exploration Survey Questions to measure sensemaking effects

After my experience with the website today:

QS1 (Reflection): I found that using the website made me reflect on the debate in a deeper way
QS2 (Insight): I found that using the website provided me with unexpected insights on the debaters and on what they said
QS3 (Focus): I found that using the website made me really focus on different aspects of the debate
QS4 (Argumentation): I found that using the website did not help me to reconstruct the arguments that the speakers made
QS5 (Explanation): I found that using the website helped me better identify and explain issues
QS6 (Evaluate Facts and Evidences): I found that using the website provided me with ways to evaluate facts and evidence
QS7 (Distinguish): I found that using the website did not help me to distinguish between the debaters' claim
QS8 (Assess Assumptions): I found that using the website did not help me to assess my assumptions
QS9 (Change Assumptions): I found that using the website changed some initialassumptions I had before the debate

References

Ackerman, B. (1989). Why Dialogue? *The Journal of Philosophy*, *86*(1), 5. http://doi.org/10.2307/2027173

Alsufiani, K., Attfield, S., & Zhang, L. (2017). Towards an instrument for measuring sensemaking and an assessment of its theoretical features. http://doi.org/10.14236/ewic/HCI2017.86

Anderson, C. W. (2011). Deliberative, Agonistic, and Algorithmic Audiences: Journalism's Vision of its Public in an Age of Audience Transparency. *International Journal of Communication*, *5*(0), 19.

Anstead, N., & Ben O'Loughlin. (2011). The Emerging Viewertariat and BBC Question Time: Television Debate and Real-Time Commenting Online. *The International Journal of Press/Politics*, *16*(4), 440–462. http://doi.org/10.1177/1940161211415519

Aubert, O., & Prié, Y. (2005). Advene: active reading through hypervideo. *the sixteenth ACM conference* (pp. 235–244). New York, New York, USA: ACM. http://doi.org/10.1145/1083356.1083405

Bansler, J. P., & Havn, E. C. (2006). Sensemaking in Technology-Use Mediation - Adapting Groupware Technology in Organizations. *Computer Supported Cooperative Work*, *15*(1), 55–91. http://doi.org/10.1007/s10606-005-9012-x

Bibiloni, T., Mascaro, M., Palmer, P., & Oliver, A. (2015). A Second-Screen Meets Hypervideo, Delivering Content Through HbbTV (pp. 131–136). Presented at the the ACM International Conference, New York, New York, USA: ACM Press.

Campo-Arias, A., & Oviedo, H. C. (n.d.). Psychometric properties of a scale: internal consistency. *Revista De Salud Pública*, *10*(5), 831–839.

César Garcia, P. S., & Geerts, D. (2016). Social Interaction Design for Online Video and Television.

Chambel, T., Zahn, C., Finke, M., 2004. (n.d.). Hypervideo design and support for contextualized learning (pp. 345–349). Presented at the IEEE International Conference on Advanced Learning Technologies, 2004. Proceedings., IEEE. http://doi.org/10.1109/ICALT.2004.1357433

Chinn, C. A., Record, R. A. T. C., 1998. (n.d.). The structure of discussions that promote reasoning. *Psycnet.Apa.org*.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences 2nd edn.

Coleman, S. (2012). Debate on Television: The Spectacle of Deliberation. *Television & New Media*, *14*(1), 20–30. http://doi.org/10.1177/1527476411433520

Coleman, S., & Moss, G. (2015). Rethinking Election Debates: What Citizens Are Entitled to Expect. *The International Journal of Press/Politics*, *21*(1), 3–24. http://doi.org/10.1177/1940161215609732

Coleman, S., Blumler, J., Moss, G., & Homer, M. (2015). The 2015 Televised Election Debates; Democracy on Demand?

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. http://doi.org/10.1007/BF02310555

Dahlgren, P. (2009). Media and political engagement. Citizens, Communication, and Democracy.Cambridge. Cambridge University Press.

De Liddo, Anna, Sándor, Agnes, Buckingham Shum, Simon, 2012. Contested Collective Intelligence - Rationale, Technologies, and a Human-Machine Annotation Study. Computer Supported Cooperative Work, 21 (4–5), 417–448. doi:http://doi.org/10.1007/s10606-011-9155-x.

Deng , A. , Dmitriev , P. , Gupta , S. , Kohavi , R. , Raff , P. , & Vermeer , L. (2017 , August). A/B Testing at Scale: Accelerating Software Innovation . In Proceedings of the 40th International ACM SIGIR Conference onResearch and Development in Information Retrieval (pp. 1395 - 1397). ACM .

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. http://doi.org/10.1111/bjop.12046

Enslin, P., Pendlebury, S., & Tjiattas, M. (2001). Deliberative Democracy, Diversity and the Challenges of Citizenship Education. *Journal of Philosophy of Education*, *35*(1), 115–130. http://doi.org/10.1111/1467-9752.00213

Faridani, S., Bitton, E., Ryokai, K., & Goldberg, K. (2010). Opinion space: a scalable tool for browsing online comments. *the 28th international conference* (pp. 1175–1184). New York, New York, USA: ACM. http://doi.org/10.1145/1753326.1753502

Fourney, A., Racz, M. Z., Ranade, G., Mobius, M., Horvitz, E., 2017. November). Geographic andTemporal Trends in Fake News Consumption During the 2016 US Presidential Election. In: Proceedingsof the 2017 ACM on Conference on Information and Knowledge Management. ACM, pp. 2071–2074.

Gergle, D., and Tan, D. S. (2014). Experimental research in HCI. In Ways of Knowing in HCI (pp. 191- 227).Springer, New York, NY.

Golbeck, J., Ashktorab, Z., Banjo, R. O., Berlinger, A., Bhagwan, S., Buntain, C., et al. (2017). A Large Labeled Corpus for Online Harassment Research. *WebSci*, 229–233. http://doi.org/10.1145/3091478.3091509

Goodin, R. E. (2003). Democratic Deliberation Within. In *Debating Deliberative Democracy* (pp. 54–79). Oxford, UK: Blackwell Publishing Ltd. http://doi.org/10.1002/9780470690734.ch3

Goodin, R. E., & Niemeyer, S. J. (2016). When Does Deliberation Begin? Internal Reflection versus Public Discussion in Deliberative Democracy:. *Political Studies*, *51*(4), 627–649. http://doi.org/10.1111/j.0032-3217.2003.00450.x

Gordon, E., & Manosevitch, E. (2010). Augmented deliberation: Merging physical and virtual interaction to engage communities in urban planning:. *New Media & Society*, *13*(1), 75–95. http://doi.org/10.1177/1461444810365315

Graham, T., & Wright, S. (2013). Discursive Equality and Everyday Talk Online: The Impact of "Superparticipants." *Journal of Computer-Mediated Communication*, *19*(3), 625–642. http://doi.org/10.1111/jcc4.12016

Grasso, A., & Convertino, G. (2012). Collective Intelligence in Organizations: Tools and Studies. *Computer Supported Cooperative Work (CSCW)*, *21*(4-5), 357–369. http://doi.org/10.1007/s10606-012-9165-3

Guess, A., Nyhan, B., Reifler, J., 2018. Selective Exposure to Misinformation: Evidence from theconsumption of fake news during the 2016 US presidential campaign. European Research Council.

Habermas, J. (1984). The theory of communicative action: Vol. 1. Reason and the rationalization of society (T. McCarthy, Trans.).

Halasz, F. G., Moran, T. P., & Trigg, R. H. (1986). Notecards in a nutshell. *ACM SIGCHI Bulletin*, *17*(SI), 45–52. http://doi.org/10.1145/30851.30859

Hancock, J. T., Landrigan, C., Silver, C., 2007. Expressing emotion in text-based communica-tion. In: Proceedings of the SIGCHI conference on human factors in computing systems(pp. 929–932). ACM, New York, NY.

Hong, L., Chi, E. H.-H., Budiu, R., Pirolli, P., & Nelson, Les. (2008). SparTag.us - a low cost tagging system for foraging of web content. *Avi*, 65. http://doi.org/10.1145/1385569.1385582

Iandoli, L., Quinto, I., De Liddo, A., & Buckingham Shum, S. (2014). Socially augmented argumentation tools: Rationale, design and evaluation of a debate dashboard. *International Journal of Human-Computer Studies*, *72*(3), 298–319. http://doi.org/10.1016/j.ijhcs.2013.08.006

Iandoli, L., Quinto, I., Spada, P., Klein, M., Calabretta, R. (2018) Supporting argumentation in online political debate: Evidence from an experiment of collective deliberation. new media & society 20 (4), 1320–1341

Kahneman, D. (2012). Ons feilbare denken: thinking, fast and slow.

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual Analytics: Definition, Process, and Challenges. In *Information Visualization* (Vol. 4950, pp. 154–175). Berlin, Heidelberg: Springer, Berlin, Heidelberg. http://doi.org/10.1007/978-3-540-70956-5_7

Klein, M., G Convertino - Journal of Social Media for Organizations, 2015. (n.d.). A roadmap for open innovation systems. *Mitre.org*.

Klein, M., Spada, P., International, R. C. P. F. 1., 2012. (n.d.). Enabling deliberations in a political party using large-scale argumentation: A preliminary report. *Academia.Edu*.

Klein, Moon, & Hoffman. (2006). Making Sense of Sensemaking 2: A Macrocognitive Model. *IEEE Intelligent Systems*, *21*(5), 88–92. http://doi.org/10.1109/MIS.2006.100

Kohavi, R., Crook, T., Longbotham, R., Frasca, B., Henne, R., Ferres, J. L., Melamed, T., 2009. Onlineexperimentation at Microsoft. Data Mining Case Studies 11.

Kohavi, R., Henne, R. M., Sommerfield, D., 2007. August). Practical guide to controlled experiments onthe web: listen to your customers not to the hippo. In Proceedings of the 13th ACM SIGKDDinternational conference on Knowledge discovery and data mining. ACM, pp. 959–967.

Kriplean, T., Bonnar, C., Borning, A., Kinney, B., & Gill, B. (2014). Integrating on-demand fact-checking with public dialogue. *the 17th ACM conference* (pp. 1188–1199). New York, New York, USA: ACM. http://doi.org/10.1145/2531602.2531677

Kriplean, T., Morgan, J., Freelon, D., Borning, A., & Bennett, L. (2012). Supporting reflective public thought with considerit (p. 265). Presented at the the ACM 2012 conference, New York, New York, USA: ACM Press. http://doi.org/10.1145/2145204.2145249

Lampinen, A., McMillan, D., Brown, B., Faraj, Z., Cambazoglu, D. N., & Virtala, C. (2017). Friendly but not Friends: Designing for Spaces Between Friendship and Unfamiliarity. *the 8th International Conference* (pp. 169–172). New York, New York, USA: ACM. http://doi.org/10.1145/3083671.3083677

Landemore, H. E., & Mercier, H. (2010). "Talking it Out": Deliberation with Others Versus Deliberation Within. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.1660695

Lawrence, J., & Mining. (2015). Combining argument mining techniques. *Aclweb.org*.

Lawrence, J., & Reed, C. (2020). Argument Mining: A Survey. *Computational Linguistics*, *45*(4), 765–818. http://doi.org/10.1162/coli_a_00364

Leiva , L. A. , and Vivó , R. , 2012 . Interactive hypervideo visualization for browsing behavior analysis . In Proceedings of the 21st International Conference on World Wide Web (pp. 381 - 384). ACM .

Lippi, M., & Torroni, P. (2015). Argument Mining: A Machine Learning Perspective. In *Theory and Applications of Formal Argumentation* (Vol. 9524, pp. 163–176). Cham: Springer, Cham. http://doi.org/10.1007/978-3-319-28460-6_10

McDonald, R. P. (2013). Test Theory: A Unified Treatment.

MacKenzie, I. S., Zhang, S. X., 1999. The design and evaluation of a high-performance soft keyboard. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, NY, pp. 25– 31.

McLeod, D. M., & Perse, E. M. (2016). Direct and Indirect Effects of Socioeconomic Status on Public Affairs Knowledge:. *Journalism Quarterly*, *71*(2), 433–442. http://doi.org/10.1177/107769909407100216

Nelson, L., Held, C., Pirolli, P., Hong, L., Schiano, D., & Chi, E. H. (2009). With a little help from my friends: examining the impact of social annotations in sensemaking tasks. *the SIGCHI conference* (pp. 1795–1798). New York, New York, USA: ACM. http://doi.org/10.1145/1518701.1518977

O'Day, V. L., & Jeffries, R. (1993). Orienteering in an information landscape: how information seekers get from here to there. *the SIGCHI conference* (pp. 438–445). New York, New York, USA: ACM. http://doi.org/10.1145/169059.169365

Ofcom, 2014. The Communications Market Report 2014. https://www.ofcom.org.uk/research-and-data/multi-sec tor-research/cmr/cmr14

Ofcom, 2019. The Communications Market Report 2019. https://www.ofcom.org.uk/research-and-data/multi-sec tor-research/cmr/cmr-2019

Olson, J. S., Olson, G. M., Storrøsten, M., Carter, M., 1993. Groupwork close up: A compari-son of the group design process with and without a simple group editor. ACM Transactions onInformation Systems 11 (4), 321– 348.

Pantazos, K., IVAPP, S. L. G., 2012. (n.d.). Constructing Visualizations with InfoVis Tools-An Evaluation from a user Perspective. *Academia.Edu.*

Pirolli, P., on, S. C. P. O. I. C., 2005. (n.d.). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. *E-Education.Psu.Edu*

Plüss, B., & De Liddo, A. (2018). Democratic Replay: Enhancing TV Election Debates with Interactive Visualisations. *Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences. http://doi.org/10.24251/HICSS.2018.215

Rainie, L., Smith, A., 2012. Politics on social networking sites. Pew Internet and American Life Project 610.

Renton, A., & Macintosh, A. (2007). Computer-Supported Argument Maps as a Policy Memory. *The Information Society*, *23*(2), 125–133. http://doi.org/10.1080/01972240701209300

Robertson, S. P., Wania, C. E., Abraham, G., & Park, S. J. (2008). Drop-Down Democracy: Internet Portal Design Influences Voters' Search Strategies (pp. 191–191). Presented at the 2008 The 41st Annual Hawaii International Conference on System Sciences, IEEE. http://doi.org/10.1109/HICSS.2008.131

Russell, D. M., CHI, M. S. S. T., 2004. (1AD). Measuring the Tools and Behaviors of Sensemaking.

Russell, D. M., Jeffries, R., CHI, L. I. S. W. A., 2008. (n.d.). Sensemaking for the rest of us. *Researchgate.Net*.

Russell, D. M., Pirolli, P., Furnas, G., Card, S. K., & Stefik, M. (2009). Sensemaking workshop CHI 2009. *the 27th international conference extended abstracts* (pp. 4751–4754). New York, New York, USA: ACM. http://doi.org/10.1145/1520340.1520732

Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K. (1993). The cost structure of sensemaking. *the SIGCHI conference* (pp. 269–276). New York, New York, USA: ACM. http://doi.org/10.1145/169059.169209

Sawhney, N., Balcom, D., & Smith, I. (1996). HyperCafe: narrative and aesthetic properties of hypervideo. *the the seventh ACM conference* (pp. 1–10). New York, New York, USA: ACM. http://doi.org/10.1145/234828.234829

Semaan, B. C., Robertson, S. P., Douglas, S., & Maruyama, M. (2014). Social media supporting political deliberation across multiple public spheres (pp. 1409–1421). Presented at the the 17th ACM conference, New York, New York, USA: ACM Press. http://doi.org/10.1145/2531602.2531605

Semaan, B., Faucett, H., Robertson, S. P., Maruyama, M., & Douglas, S. (2015). Designing Political Deliberation Environments to Support Interactions in the Public Sphere. *the 33rd Annual ACM Conference* (pp. 3167–3176). New York, New York, USA: ACM. http://doi.org/10.1145/2702123.2702403

Sereno, B., Shum, S. B., & Motta, E. (2005). ClaimSpotter: an environment to support sensemaking with knowledge triples. *the 10th international conference* (pp. 199–206). New York, New York, USA: ACM. http://doi.org/10.1145/1040830.1040875

Shao, C., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2016). Hoaxy: A Platform for Tracking Online Misinformation. *the 25th International Conference Companion* (pp. 745–750). New York, New York, USA: International World Wide Web Conferences Steering Committee. http://doi.org/10.1145/2872518.2890098

Shipman, F. M., Marshall, C. C., 1999. Formality considered harmful: Experiences, emerging themes, and directions on the use of formal representations in interactive systems. Computer Supported Cooperative Work 8 (4), 333–352.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22–36. http://doi.org/10.1145/3137597.3137600

Snell, P. (2010). Emerging Adult Civic and Political Disengagement: A Longitudinal Analysis of Lack of Involvement With Politics:. *Journal of Adolescent Research*, *25*(2), 258–287. http://doi.org/10.1177/0743558409357238

Spink, A., Wilson, T. D., Ford, N., Foster, A., & Ellis, D. (2002). Information seeking and mediated searching study. Part 3. Successive searching. *Journal of the American Society for Information Science and Technology*, *53*(9), 716–727. http://doi.org/10.1002/asi.10083

STOKER, G., HAY, C., & BARR, M. (2016). Fast thinking: Implications for democratic politics. *European Journal of Political Research*, *55*(1), 3–21. http://doi.org/10.1111/1475-6765.12113

Susskind, J. (2018). Future Politics.

Sunstein, C. R., 2018. # Republic: Divided democracy in the age of social media. Princeton University Press.

Taber, C. S., & Lodge, M. (2006). Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*, *50*(3), 755–769. http://doi.org/10.1111/j.1540-5907.2006.00214.x

Towne, W. B., Technology, J. H. J. O. I., 2012. (n.d.). Design considerations for online deliberation systems. *Taylor & Francis*.

Van Deventer, M. O., de Wit, J. J., Guelbahar, M., Cheng, B., Marmol, F. G., Köbel, C., et al. (2013). Towards Next Generation Hybrid Broadcast Broadband, Results from FP7 and HBBTV 2.0. *International Broadcasting Convention (IBC) 2013 Conference*, 12.3–12.3. http://doi.org/10.1049/ibc.2013.0049

Weick, K. E. (1995). Sensemaking in Organizations. SAGE.

Weinmann, C. (2017). Measuring Political Thinking: Development and Validation of a Scale for “Deliberation Within.” *Political Psychology*, *39*(2), 365–380. http://doi.org/10.1111/pops.12423

Wilson, M. J., & Wilson, M. L. (2013). A comparison of techniques for measuring sensemaking and learning within participant-generated summaries. *Journal of the American Society for Information Science and Technology*, *64*(2), 291–306. http://doi.org/10.1002/asi.22758