Deliberation, argumentation, and democracy

Paper prepared for APSA Annual Meeting, Montreal, 2022

Mestre R., Ryan, M. Gheasi, M. et al. (University of Southampton)

Work in progress – Comments welcome.

Correspondence – Dr Matt Ryan, M.g.ryan@soton.ac.uk, Dr Rafael Mestre R.Mestre@soton.ac.uk

Advanced democracies face a plethora of wicked problems of governance linked to increased polarisation of politics, the spread of misinformation, and decreased trust in democratic institutions. One common point of contact in all these issues is deliberation and/or argumentation, both online and offline. Computer scientists and philosophers of linguistics have tried to understand how arguments can be identified, their quality assessed, mapped, and improved, with techniques ranging from artificial intelligence and machine learning to qualitative methods. Democratic Innovations, inspired by theories of deliberative and participatory democracy have focused on institutional engineering to increase inclusion and capacity of citizen voices. This paper maps the different foci of responses to these problems and identifies missed opportunities for clever collective response. We show that conceptual confusion as well as differences in focus on argumentation, dialogue or discourse has led to underuse of deliberative insights in argument mapping, in turn reducing the impact of advances in argument identification in democratic innovation. The paper discusses implications for some of the leading tools and platforms currently in widespread use. We find that approaches in the social sciences have advanced strong normative criteria, as well as detailed policy implication, but lack a mid-range theory to explain how affordances of design affect communication in fora. Elsewhere certain approaches linking computer science and philosophy have offered strong conceptual theory to guide efficient product design but suffer from a lack of attention to normative questions of how social outcomes are achieved. Drawing on these insights we produce a novel organising perspective to guide efficient discovery of solutions within this pressing research agenda.
1. Introduction

If deliberation, as a means of releasing the force of the better argument is a necessary condition for the legitimation of democracy, how do we understand and engender it? Much work has been done to impose discipline on the study of democratic deliberation. This disciplinary work provides an important shared organising perspective for discoveries but can also limit cross-fertilisation and sustain lacunae by encouraging homophilic scholarship. In this paper we provide an original account of the different disciplinary perspectives that have produced significant bodies of work on deliberation and argumentation in democratic governance. We provide a description of their magnitude and core features, including key developments and trajectories. Our aim is to provide a rigorous assessment of different approaches to studying a phenomenon of recognised importance for democrats. By way of critique we identify lacunae within certain disciplinary approaches and offer ideas to help scholars better traverse these fields to conceptualise, observe and analyse democratic deliberation in modern fora.

Across the turn of the 21st century, a significant body of work in the tradition of the policy sciences called for an ‘argumentative turn’ in policy analysis (Dryzek 1990, Fischer and Forrester 1993, Hajer and Wagenaar 2003, Fischer and Gottweis 2012). Most contributors to these debates positioned themselves as critical interpretivists, looking to rescue understandings of democratic politics from a technocratic and narrowly positivistic version of elite-driven policy. Primarily they sought a greater role for reason-making as a process of communication among plural actors lying at one remove from observation and testing of scientific process. A logical entailment of such arguments was for greater inclusion of diverse voices in policy deliberation, though such participation plays varying roles across accounts of deliberative democracy. Ironically, almost in parallel to the deliberative turn in democratic theory and its establishment as a key concern within political science (Spada and Ryan 2017), with the acceleration in artificial intelligence, collaborations between philosophers and computational scientists began to focus more seriously on the scientific categorisation and extraction of arguments according to more fixed schema or frameworks. Although the term ‘deliberation’ bears important normative assumptions in political theory that marry participatory commitments to understanding of communicative rationality, it is perhaps the familiarity with this term over the term argumentation that has led to little cross-fertilisation with other fields of applied argumentation, such as has happened in computer science and education.

Introductions to the interdisciplinary field of argumentation commonly refer to the work of argumentation theorists like Doug Walton and Erik Krabbe. They present a classification of human dialogue based around the varying objectives of the participants and the information that was available for each participant at the beginning of the discussion (McBurney et al., 2007). Six primary types of dialogue have been defined in Walton and Krabbe’s scheme: persuasion dialogues, information seeking dialogues, inquiry dialogues, negotiation dialogues, eristic dialogues and, finally, deliberation dialogues. The first five types have been widely studied in the literature, to different extents. Formal models to describe them have been proposed, even including combined models that try to better represent real-life dialogues, in which several of these ideal types overlap (McBurney et al., 2007). Deliberation dialogues, of special interest for social scientists, where participants jointly decide the course of action to solve a particular issue, is perhaps the least studied by formal argumentation modellers.

Beyond theories and models, technology-enabled argumentation tools have been increasingly applied in the last 30 years in research, but also by practitioners in politics, education and law. For
instance, in the United States, governmental bodies have started to apply initiatives of electronic rulemaking (eRulemaking) to assist in the decision- and rule-making process. At the University of Cornell between 2005 and 2017 a multidisciplinary group called CeRI (Cornell e-Rulemaking) was created to “[assess] the value of technology-enabled rulemaking participation and offer specific principles of participation-system design”, with emphasis on the introduction of human moderators. Argumentation models that adapt to legal decision making have been proposed (Park et al., 2015). Researchers in education have developed tools to help students learn important argumentation skills through “collaborative argumentation” – experiential learning by debating topics from a scientific or legal point of view (Scheuer et al., 2010).

In politics, democratic innovations have taken the form of participatory budgeting, referenda, citizens’ assemblies and other mini-publics, or, co-governance and co-design (Elstub and Escobar 2019). These innovations in institutional design have taken considerable inspiration from theories of deliberation as a vehicle for engagement, learning and discussion or argumentation. Deliberative democracy is a normative theory that aims to put better (read both more legitimate and more intelligent) collective decision-making at the centre of policy-making (Smith 2009).

Yet in practice these developments in theory and innovation have yet to fully land any real transformation of mass political communication. If anything the transformation of participation in the public sphere has been dominated by privately-owned social media platforms that have been characterised as often less legitimate media for political debate and are hardly consensus-oriented. The participation of non-expert citizens in legal or governmental issues has increased the number and diversity of arguments that transmit from public spheres to decision-making, “whose validity and strength are difficult to evaluate, both by the government agencies and fellow citizens” (Park et al., 2015). Deliberation promises an inclusion that overcomes pathologies of a democracy without strong opportunities for free communication and interpretation. But it must overcome the challenge of information overload that incentivises a reduction back to simple aggregation of preferences conceived by narrow agendas. Computational Argumentation tools have been presented to assist deliberation in political settings (Iandoli et al., 2018; Klein et al., 2012), and may show some promise for future of democratic debate but the question of how to sequence these devices and how best they might be adopted in the public sphere remains open. In what follows we first briefly describe the contributions and lacunae in efforts to design online debate, especially with the advent of machine learning applications in natural language. We then proceed to chronicle and better describe than has been done before some of the different foci that research groups have brought to the goal of improving deliberation in the online public sphere. We continue by elaborating some of the more successful contributions of philosophers and computer engineers working in these subfields, before explaining the distance between current models for representing political speech and potential. We conclude with a discussion of potential next steps for tools that aim to support collective reasoning.

2. Designing argumentation in the Dark? Dispersion of Theories, Tools and Measures

In the last 30 years, social science researchers and practitioners have developed a wide variety of tools to support decision-making, deliberation, and cooperation both online and offline. For instance, ParticipateDB1 collects a list of different types of platforms that can be used to assist deliberation and participation in political or business settings, from argumentation mining to online

1 http://www.participatedb.com/categorizations
deliberation and social media. In other fields like education, argumentation tools can support learning in classrooms by enabling students to obtain argumentation skills, with Scheuer offering a good review of the existing argumentation tools for that purpose already used by 2010 (Scheuer et al., 2010). Many of the tools that have been developed, like the Parmenides system (Atkinson et al., 2004, 2005), the HERMES system (Karacapilidis & Papadias, 2001), the Demos system (Richter & Gordon, 2002), Carneades (Gordon et al., 2007a), Araucaria (Reed & Rowe, 2004), Belvedere2 (Suthers, 2014), LARGO (Pinkwart et al., 2006), or the Deliberatorium (Klein et al., 2012), have been used in real life settings to promote or assist deliberation in political or educational landscapes (Scheuer et al., 2010; Schneider et al., 2013). However, as Scheuer et al. recognised in their survey (Scheuer et al., 2010), there are countless different design options in terms of visualisation of the argument, the interaction with the machine, and provision of machine-assisted feedback. This calls for “a more general theoretical framework about argumentation and its support with technology”, since generally “there are no two systems that have the same ontology” and “often, different names are used for the same basic concepts” (Scheuer et al., 2010).

The plethora of argumentation tools and platforms with different design decisions and theoretical backgrounds has become overwhelming and is only expected to increase with the inclusion of machine learning technologies. The cross-talk between social science and computer science was, up to approximately 10 years ago, in a stable position. Most colleagues understood that the other had to exist somewhere else on campus, and a few collaborated - as computer scientists contributed with software development and visualisation tools for social science research. Since then, more researchers, practitioners and policy makers are becoming aware of the potential of artificial intelligence and machine learning technologies. In fact, computational social science labs have become a feature of the landscape. However, the superficial and actual complexity of AI creates a significant wall that prohibits non-expert researchers from interrogating advances in platforms, research, and interventions even they themselves have designed, without collaboration from machine learning experts, often speaking a completely different language (Parsons 2019).

There is a great potential for these technologies to provide an improved machine-assisted deliberation, for instance collecting and summarising arguments, providing feedback to the user, or retrieving further information from the web about a certain topic of discussion to educate the debater. The tools to enable such are already available or are currently in a late stage of development. Yet no argumentation tools have fully taken advantage of these advances. This has created a division in the literature (if there was not one to begin with). A rapidly increasing number of machine learning and neural networks approaches to argumentation mining, and natural language processing (NLP) in general, try to solve loosely developed research questions borrowed from the social sciences and are published in specialised outlets, outside of the view and the understanding of researchers and practitioners that would find them interesting.

In multidisciplinary collaboration, philosophers of argumentation, social scientists from different disciplines (from education to politics and law), computer scientists and practitioners of democracy might better discuss important issues that often are left to one or two researcher’s individual decisions, such as visualisation, ontologies, and user interaction. There might never be a consensus around which argumentation framework is the most appropriate or can efficiently adapt to different scenarios, but we might gain better understanding of a range of tools and their appropriate use. We argue that a functional and replicable tool, although perhaps not completely efficient, can be more useful to enable machine-assisted deliberation in real-life settings than a

---

http://belvedere.sourceforge.net/
multiplicity of theoretically strong but uni-disciplinary and barely tested tools that happen to miss many in the field they want to reach due to lack of collaboration.

### 2.1 Tools for computer-assisted argumentation

Given the large scale of deliberations occurring online in social networks, forums or blogs, social scientists have investigated the use of custom-made argumentation systems to enable these deliberations in a way that can be useful, both for the people using them and for decision makers if they are to be used in group decision events. These tools offer different collaborative environments, that vary in the degree to which they constrain the user, and promote more or less a users’ horizontal interaction with one another. At one end of the spectrum, we have fully flexible collaborative argument mapping tools that allow users to propose ideas, raise issues with such ideas, identify arguments in favour or against a proposition. Such collaborative tools generate maps with a deep topology that are appropriate for complex ideation and deliberation (example: Deliberatorium). At the immediate level we have platforms that focus primarily on providing arguments in favour or against a set of predefined options and are more appropriate for debate-style discussions (example: Kialo, Consider.it). Lastly Pol.is is a platform that generates a map of preferences regarding open comments added by organizers and participants. It leverages a voting mechanism, a process that is particularly useful for supporting temperature-check exercises and the initial mapping of ideas in a community. The level of horizontal interactions among users in these argument mapping tools is limited by the possibilities generated by the map. In the Deliberatorium participants can raise issues and offer pros and cons to entries provided by other users generating a sort of mediated dialogue. In Kialo and Consider.it the users can offer pros and cons. Lastly in pol.is the users can only offer a like or a dislike or a pass to other participants’ comments.

These tools purport to provide a necessary upgrade in affordances to status quo political deliberation. One of the criticisms of classical deliberation technologies, like forums or chat rooms, is that they are time-centric and conversation-centric and, therefore, their content is scattered with high levels of noise (Klein et al., 2012), making it difficult to harvest the most important conclusions, ideas, or arguments. Topic-centric tools like wikis, on the other hand, do not suffer from content scattering of low signal-to-noise ratio, but they do encounter other problems like redundancy. Argument-centric tools, like Computer-Supported argument visualisation (CCSAV) tools, have been thus proposed to solve many of these issues. These representation-centric platforms deliberately suppress elements from typical face to face deliberation, such as turn taking, or the reply structure of online deliberation in forums, to focus on the arguments being made (Iandoli et al., 2016). In this framework of deliberation, arguments are shown to the users in an argument map, each node representing a single idea (reducing redundancy and increasing signal-to-noise ratio) with links to other nodes in the form of for/against or support/attack relationships that remain close to one another (reducing content scattering). One of the most successful examples of this has been the Deliberatorium (Klein et al., 2012). In a real-case use of this platform, users were recruited from an online group of the Italian Democratic Party to deliberate about the reform of Italian electoral law. With an age distribution comparable to that of the country, the users were divided into different groups comparing debate in a typical time-centric forum with using argument maps. This research evidenced that the exposition of ideas in forums tended to be more redundant and self-referential, although the engagement of users with the CCSAV platform and the overall production of ideas was smaller, possibly related to a lack of familiarity with a tool unlike a forum. Likewise, similar studies have pointed to the perceived negative quality of collaboration in argument-centric platforms when compared to forums (Iandoli et al., 2016).
The emerging research on these argument mapping platforms hints at a supposed trade-off between user-friendliness and the impact of the deliberation, which perhaps reflects a familiar conundrum of political participation and education or size and democracy more broadly. While users might perceive more difficulty using an argument-centric tool, redundancy and content scattering could become problems for aggregation of the ideas and arguments, requiring extensive human intervention to manually aggregate them. This decreases potential impact in use-cases in distributed decision-making, collective deliberation, and other forms of democratic innovations (Iandoli et al., 2018). Despite joint efforts of academics in Web science, social science, and political science to design and develop these tools, their wide usage is still far away, as policy-makers seem mostly unaware of these tools, and most political debate continues to take place on platforms which were not designed with political debate in mind. In contrast, social media tools were mostly designed to enable expression and advertising as well as developing profiles of users.

Time consumption and the learning curve of a new technology and its implications is a difficult wall to surpass, especially in a landscape were political interests and decision-making processes shift rapidly. At their current state, more specialised platforms require a moderator or academic researcher to ensure the validity of the arguments and ensure the correct use of these platforms. Automatization and summarisation tools might help decrease this burden, and interdisciplinary collaboration with computer scientists might be speed up this process. At the moment design continues to fail to meet demand (at least in terms of demand for enhanced political deliberation in the public sphere). In the next section we further explain and identify missed opportunities.

2.2. Same goals, different foci? Identifying and measuring deliberation/argumentation

Previously distant disciplines like political science, computer science or arms of philosophy have focused their efforts on different aspects of dialogical deliberation and argumentation, often creating their own ontologies, rules, and tools to assist in their research or applications, unaware that they occasionally share the same goals with colleagues in other disciplines. That is not a new lament but manifests in particular problems for developing norms for deliberation mediated through applications based on the web, and employing AI.

Political science in its assessment of political communication has given an understandable emphasis to the study of political debate among political elites, where debates are held publicly following rules and customs, and most generally take the form of an ordered tree-like structure of a common dialogue. Within dialogical structures, we could consider traditional political debates a distinct substructure, in which participants generally perform a long speech, which can usually be refuted or commented upon by the next participant, with strict rules of behaviour to ensure an efficient debate and that all participants have time to present their opinions. Computer scientists, on the other hand, have been more interested in a classical substructure, that we could call a “pure dialogue”, in which the interventions between speakers are shorter and the responses are immediate, rather than delayed until the other person has finished their speech.

Philosophy of argumentation has placed emphasis on the evaluation of what constitutes a good argument and has sometimes inspired operationalisation. Normative principles of deliberation are useful metrics to validate the outputs of argumentation frameworks or dialogue models. One theory that has inspired much work across broad scholarship in aspects of debate style communication is Jurgen Habermas’ discourse ethics, which gave rise to the Discourse Quality Index (DQI), a score intended to code and measure the quality of a discourse, based on Habermas’ theory (Steenbergen et
This index has found numerous applications in work on democratic innovations, as a way to measure and compare different deliberative platforms like mini-publics or citizens’ assemblies to help design events and evidence their quality (Escobar & Elstub, 2019). The DQI’s coding categories ought to be easily scalable and generalisable for further analysis by computers. One criticism, however, remains that the analysis can require too much normative content for observational comparison, as it relies on justifications of the common good in utilitarian terms or the difference principle (Steenbergen et al., 2003). Moreover despite strong interrater reliability in general, annotation of DQI has generally been performed up to now by users with strong training in deliberative norms. Automation or crowdsourcing have not been reliably used with these measures.

Two other sets of rules are Robert Alexy’s rules for discourse ethics., also based on Habermasian philosophy, and Hitchcock’s Principles of Rational Mutual Inquiry, aimed at ensuring a rational agreement between participants in a discourse. For the sake of conciseness, we will not reproduce the rules here, but they can be checked elsewhere (McBurney et al., 2007). Both sets of rules offer an interesting example of how interdisciplinary research can improve assessment of deliberative processes against some normative theory. Again some rules and rule-breaking are easier to identify than others. Whereas rules of contradiction or justification can be assessed more easily through NLP and argumentation mining to understand how people deliberate, more complex rules that assert freedom of expression without constraint require more input from social theorists. Some other rules like those requiring a speaker to be truthful to themselves are virtually impossible to observe, but perhaps further research in rhetorical aspects of discourse like ethos and background knowledge could bring computers closer to a level of judgement that is comparable, or even better, than that of humans on these or related attributes of debate such as probabilistic measure of ‘truth to oneself’.

One of the most straightforward examples of collaboration that has occurred between disciplines has been the application of machine learning or artificial intelligence techniques to analyse and understand deliberative political debates. However, this poses some issues for machine learning-based deliberation frameworks. Recent advances in NLP, such as recursive neural networks like long short-term memory arrays (Schuster & Paliwal, 1997) or transformer architectures (Vaswani et al., n.d.), have allowed algorithms to pay attention to specific parts of a sentence and retrieve some memory of what has been said in order to improve the analysis of argumentative features. However, computers still struggle to realise when a word or an entity has stopped having importance in a text and the topic has shifted. Most of the approaches to classifying relationships between arguments in a dialogue (e.g. support, attack, rebuttal, warrant…) are usually based on comparing pairs of sentences that can be of two types: i) close-by sentences uttered by different people engaged in a dialogue; ii) two sentences uttered by the same person during a deliberation. In both cases, the closeness of the sentences is a feature that classifies whether the sentences are strongly related or not. In political dialogue, this approach may be error-strewn. During a long speech, a politician might make reference to different parts of the speech of the preceding person, something that a computer might have problems understanding in the usual approach. In fact, considering what someone has said a while previously is taken as a very good sign of reciprocity in observations of deliberation.

Finally, more complex applications involve discourse mediation by multi-agent systems (MAS) which could enforce rules of respect between speakers and requirements of justification of arguments. Whether this type of normative mediation can improve the quality of a deliberation or instead decrease its quality and freedom of expression will be an ethical matter for future research. Recent work on ‘AI facilitation’ has focused on using simple prompts for example to remind participants in a debate not to hog speaking time. But if we can agree on what makes an argument
valid, machines ought to be able to weigh the validity of arguments. That’s a big ‘if’ and perhaps ultimate agreement is not preferable (Mouffe 1999). Yet a priori agreement could lead machines to reduce power imbalances by helping otherwise disempowered actors develop arguments that better communicate for them.

Argumentation frameworks applied in MAS and similar topics should at least have an understanding of the normative aspects of argumentation or discourse ethics to adjust their behaviour to the dynamism of the deliberation. It remains to be seen if future advances in NLP could help in these validation tasks by automatising them at least in a semi-supervised manner. These systems should be heavily tested against coded bias due to the ethical concerns that indiscriminate use of ML technologies has recently brought up, in particular in computer vision applications. But NLP techniques are neither free of these concerns, as typical representations of texts like word embeddings have been shown to participate in direct gender bias across different corpora, even seemingly innocuous ones (Babaeianjelodar et al., 2020). Not only gender, but also racial bias, has been evaluated in specific tasks like sentiment analysis, providing an “equity evaluation corpora” for researchers to check whether a sentiment analysis model is scoring a specific race or gender with higher sentiment intensity (Kiritchenko & Mohammad, 2018). It is more difficult, however, to identify unintentional bias in argumentation mining tasks. Such biases would stem from at least two different sources: i) from the argument itself, by assigning higher intensity of support or attack, or argument strength, to an argument by example or expert opinion coming from specific gender or minority group; or ii) from the speaker, if the model is implemented with knowledge of personal characteristics of the speaker. Although one can argue that a symmetrical weight of influence for each participant in tasks like sentiment analysis is desirable, this might not be the case for argumentation mining tasks, in which arguments might be built around discrimination debates and premises rooted in a history of discrimination against certain social groups, where equity requires some extra weighting or redistribution of power for minorities. A human person might assign a very low argumentative strength to a premise regarding discrimination against men, but a high strength when it is applied to women. Any automated argumentation mining model would need to be aware of these differences and, although we are not aware of any work or dataset tackling the issue of bias in argumentation mining as the field is still in its infancy, we believe communication between ethicists, computer scientists and social scientists is paramount to investigate these issues before their implementation becomes more widespread. This is especially true as automation and argumentation mining move beyond applications in structured elite deliberation to wider public debates on private web-based applications.

In summary, while opportunity abounds, so do risks, and a lack of understanding across domains of expertise appears to heighten risk while reducing opportunities. Even quite simplistic applications of machine processing could greatly enhance democratic speech. There is research scattered about the field of argumentation mining focused on the extraction of arguments and their relations between them. Out of all the arguments detected in an application of an argument-centric platform, a summarisation tool might reduce redundancy, as well as collect and order all the arguments that might support or attack the underlying claim. This could be easily presented to policy-or decision-makers with simple and direct visualisations, decreasing the time involved in understanding whether the arguments are redundant or unrelated, and focusing only on understanding on what people have to say about specific topics. Of course redundant or defeated arguments do not make a person’s claim to recognition of their argument redundant, and this is an important issue to confront for those who search for some form of computer-aided rationalisation of collective decisions. For this to happen, however, all the fields at play need to realise that they are
moving towards the same goal with different foci and, on many occasions, completely different terminologies for the same concept.

2.3 Interdisciplinary collaboration

There are, naturally, promising examples of cross-boundary applications of computer science, particularly NLP, in other fields, closely related to argumentation and deliberation, that should also be considered, like for instance the detection of deception, controversy, misinformation or other types of “bad argumentation”. The detection of deception or fallacies has a strong connection with philosophy of argumentation that, although usually concerned with the normative aspects of deliberation, has recently focused on non-normative aspects, like the detection of fallacies instead of other aspects of the validity of the arguments provided (Amossy, 2017). In the computer science realm, (Kopev et al., 2019) attempted to detect deception occurring in political debates by using machine learning techniques based on analysis of audio and textual features, presenting the first multimodal database for this purpose. Deception recognition has also found a lot of interest in the detection of so-called “fake news”, both on social media and classical media outlets. Especially with the rise of social networks and collaborative platforms, manual fact-checking becomes a tedious and impossible task. There is considerable research in developing automatic tools that can, perhaps with minimal human assistance detect fallacies. Some works have attempted to capture the style of writing of fake news articles in an automatic manner, for instance with bag of words approaches - they cannot actually identify why the presumed fact in question is actually fake (Granik & Mesyura, 2017). Other works, like (Pan et al., 2018) have used incomplete knowledge graphs and news articles from trusted sources to train a fake news detector based on content. Specific social platforms, notably Twitter, have received increasing attention, as the spread of fake news has grown alongside the increase of social bots in these types of platforms (Shao et al., 2018). In this regard, (Davis et al., 2016) presented BotOrNot, a system to evaluate whether a Twitter profile can be a bot or a real person, with potential applications in fake news detection on this platform. Likewise, (Burnap & Williams, 2015) applied machine learning and statistical modelling to identify cyber hate speech surrounding live controversial events to assist human decision making for policy makers. ‘Controversy’, closely aligned to misinformation and deception, is a more classical topic of research in computational argumentation (Lawrence & Reed, 2019). For instance, (Choi et al., 2010) attempted to identify controversy from news articles based on the assumption that a controversial issue would receive strong sentimental features and (Rumshisky et al., 2017) used a network-based model to analyse connections in real time controversies, as well as word meaning drift over time, indicating diverging semantic representations. These works represent steps forward. The application of argument, controversy or misinformation detection to inform policy makers and the wider public is becoming possible, but a lack of interdisciplinary collaboration is stopping the broader implementation of these methodologies. The puzzle is why such collaboration is not happening and what stymies it.

We can take some ideas from where collaboration has worked well. When discussing interdisciplinary collaboration, John L. Pollock’s inference graphs are regarded as one of the first attempts to use AI techniques to model arguments, bringing together philosophy and computer science. As Pollock identified in his famous paper on defeasible reasoning (Pollock, 1987), “the researcher in AI can learn from the philosopher” and “the philosopher can learn much from attempts to implement epistemological theories in concrete programs for machine reasoning”. This collaboration helps us understand a little better how we might change our minds as the result of information processing in a debate. Indeed, the so-called nonmonotonic reasoning in AI and defeasible reasoning in philosophy can be argued to be the same thing. Pollock starts by assuming
that human reasoning is constructed by joining together individual inferences to form complex arguments (Pollock, 2009); in the form of inference graphs that do not necessarily form a linear sequence. In Pollock’s system, the epistemic state of an agent is represented by an inference graph, in which each node represents reasons and propositions of an argument and directed edges represent relations of support or defeat. Assuming defeasible reasoning highlights the importance of perception and memory (Pollock, 1987), but also the fact that belief formation is dynamic and beliefs can be withdrawn when further information is presented (Pollock, 2001), which returns us to the consideration of the importance of rhetorical aspects of argumentation. The existence of defeasible reasons, moreover, requires the existence of a type of “reason” called defeater, which can act on the conclusion (rebutter defeater) or on the connection between reason and conclusion (undercutting defeater). (Dung, 1995) also explored the ways in which human argumentation can be implemented in computers, including defeasible arguments. More recently, the Carneades model is an example of a defeasible framework in which opposite arguments and exceptions in this model can represent Pollock’s defeasible reasons (Gordon et al. 2007b).

(Walton, 2011) identifies a type of defeasible fallacious argument as one that despite its weakness, “is so powerfully impressive to the given audience at a particular moment, it has a devastating impact and carries the day”. Ad hominem arguments, he continues, in which a person attacks the ethos of an adversary, instead of the argument itself, are common in political discourses and might be “taken for a much stronger argument than it really is”. Likewise, fake news are aspects of another type of argument that, despite being defeated easily by rebuttal, are effective in people’s minds and whose reasons constitute a stronger argument within their belief system. Defeasible system approaches are born from simpler argumentation approaches which usually assume a “relational” approach in which “a statement ‘follows’ from a set of statements if an argument for it can be constructed that survives all attacks by counterarguments that can be constructed from the same set of statements” (Gordon et al., 2007b). In this sense, Carneades was presented as an open integration framework thought to support persuasion dialogues and is based on Walton’s philosophy of argumentation. One criticism, however, to defeasible reasoning systems like that proposed by Pollock and others, is that they still tend to consider an ideal reasoning agent behind their structure and to rely on strong logical assumptions. Therefore, in these cases, the weakness of the argument (even with internal contradictions) is not a real problem, but the ability to model the rhetorical effect is (Walton, 2011).

One of the defining characteristics of deliberation dialogue as defined in Walton and Krabbe’s formalism is that they are concerned with what should be done in a specific situation to decide on a course of action, while information or inquiry-seeking dialogues are concerned with facts or the search for truth (McBurney et al., 2007). When compared to negotiation and persuasion dialogues, we can identify concern for action as a key similarity, but in a deliberation dialogue the course of action needs to be decided jointly, and one of the participants does not necessarily need to attempt to persuade the rest to change their mind, since all of them should share a common goal (McBurney et al., 2007). However, this assumption of the common goal might not always apply in a real-life deliberative discussion and, in any case, this does not entail a common understanding between participants. Political theorists have debated whether the notion of common good (and possibly common goals) is a myth, and if so a useful one, since at least the work of the early social contract theorists. Some works in argumentation have made more effort to critique the notion of a common goal and have attempted to extend current models of deliberation to include cases like lack of common knowledge, change of circumstances, or when to move to closure (Walton et al., 2016). The consideration of what should be done in a specific situation, even if it is not the ideal course of action for individual participants, but the optimal one for the group as a whole, is one of the defining features of this type of dialogue, which
makes the number of normative theories built around it unsurprising. Also, pros and cons in arguments can easily rely on ethical aspects, for instance in debates about legalization of euthanasia, which gives them an extra normative component in their definition. Can computers learn how to account for this? New ways of assessing the ethical aspects of machine-assisted deliberation are needed, as normative theories such as Robert Alexy’s rules for discourse ethics or the DQI discussed above do not consider a scenario in which not only the participants are involved, but also machines with their own coded bias.

In these cases then, philosophy on the question of how we reason collectively and individually informs and learns from the discipline of modelling for machines to interpret. And vice versa up to a point. Yet collaborations that link theories, measures and tools remain elusive. An increasing part of the problem with the modelling of argumentation and deliberation in computer science literature is the explicability of the results. Machine learning models based on complex neural networks, have often been criticised for the impenetrability of explaining how they obtained their results, which can lead to serious ethical problems in their implementation. Moreover, in argumentation mining in particular, the increasing complexity of argument maps calls for new strategies to efficiently support real-life applications. One way to circumvent these issues is the use of symbolic or sub-symbolic approaches, as proposed by (Galassi et al., 2020), that is, approaches that are based in some way on the explicit manipulation of symbols (Sarker et al., 2021). The authors suggest the adoption of Neural Symbolic (NeSy) or Statistical Relational Learning (SRL) that enable AM frameworks to express uncertainties at the conceptual level. Instead of implementing handcrafted rules, such as those defined by certain argumentation frameworks, (sub-)symbolic approaches can enforce constraints during training, being able to learn the probabilities or rules of the framework. The authors exemplify the application of defeasible rules in the context of a political debate using prior knowledge of the probabilities that participants would attack each other’s claims. Thus, defeasible frameworks, like those described in the previous paragraphs, that represent complex human reasoning, can potentially be addressed with symbolic and sub-symbolic approaches, benefiting from their ability to adapt their rules or probabilities with new information is available. Here we see again the beginnings of attempts to marry information about different aspects of debate in identifying or predicting outcomes. Whether such attempts are adopted my stand on fall on whether participants in a debate or decision-makers are comfortable with probabilistic reasoning. Debaters with strongly held identities may not find such approaches as psychologically satisfying as deterministic ones. Symbolic approaches might need specialised and individual design rules, something which works against widespread implementation in different scenarios.

4. Where to? Opportunities for bringing rhetoric back into machine assessments of political speech

Rhetorical models of argumentation, as defined by (Bentahar et al., 2010), encompass those models that do not take into account the micro or macro structure of arguments, but their rhetorical structure as units of discourse. Rhetorical frameworks, therefore, consider also the characteristic appeal to the audience of persuasive discourses. For rhetorical theories of argumentation, the internal structure of the argument is not the main interest (and it can be completely absent), but the rhetorical aspect, namely the perspective of the audience and the ethos of the speaker, are of high importance. Thus, rhetorical arguments do not try to establish the truth of a proposition, but to consider value judgements of the arguments and have more affinities with work in political psychology. We can draw the development of rhetorical argumentation theories back to the work of Aristotle’s Rhetoric (Cope
A more recent influential text is Perelman and Olbrechts-Tyteca’s *New Rhetoric* (Perelman & Olbrechts-Tyteca, 1969). In the latter, the two authors consider that the goal of any argument is to influence or convince an audience, and therefore the techniques used during argumentation should be designed from the audience’s perspective and validated according to their effectiveness in fulfilling their goal (van Eemeren et al., 2014). In this way, Perelman and Olbrechts-Tyteca refuse the use of formal logics, as they consider them irrelevant to the study of argumentation (van Eemeren et al., 2014). They develop a taxonomy of arguments by dividing them into three classes: quasi-logical arguments, argumentation based on the structure of reality, and argumentation constructing the structure of reality. It is interesting to notice that despite their refusal to use formal logic, they admit that arguments usually draw from logical or mathematical structures to *claim to have* a logical structure. In argumentation based on the structure of reality, far from providing an ontological perspective of nature, the authors refer to the way in which arguments are constructed according to the audience’s view of reality. Likewise, an argumentation framework establishing the structure of reality uses examples, illustrations or analogies to present a reality to the audience that is new.

Needless to say, building argumentation frameworks based on rhetorical theories like that of Perelman and Olbrechts-Tyteca or similar ones that consider the audience’s perspectives can be rather challenging. Making use of knowledge requires a balance between parsimony and complexity and, ontology notwithstanding, strong claims regarding the irrelevance of formal logical for the study of argumentation are challenged by recent advances in artificial intelligence (see van Eemeren). Nevertheless, this theory places the importance of the audience when building argumentation frameworks and their applications in the spotlight. How can tools we design help us understand the extent and consequence of differing views of reality of the audience from that of the speaker or the interacting agent? We have to model multiple perspectives. A human speaker can certainly understand and adapt discourse in making an argument, emphasising the persuasiveness component of speech. However, machines as currently designed might have a more factual view of reality and find it hard to understand the audience’s view of reality and adapt their arguments to it.

Rhetorical aspects like these ones are of particular importance for two completely different branches of knowledge and technology, such as political science and multi-agent systems. MAS introduce above can choose upon a course of action by reasoning according to the available information and knowledge about the world (i.e., their environment, including other interacting agents and external inputs). Programming schemes that try to model cognitive abilities and reasoning processes for the development of MAS can include information like beliefs or desires. One of the most influential programming models for MSA is the Belief-Desire-Intention model (BDI), which can aid in decision making by considering the beliefs of the agent about its environment, the desired outcomes of the agent and the intention or series of steps needed to achieve that desire (Cardoso & Ferrando, 2021). Politicians are known to employ emotional speeches (Osnabrüge et al., 2021) or populist rhetoric (Nai & Maier, 2018) when speaking for different audiences and policy makers might want to apply policies that adjust to the views of the people they represent, therefore MAS might show promise for a more accurate modelling of political argument.

The *pathos*, or audience’s emotions, might not be the only rhetorical aspect of argumentation that machines can have problems understanding. The study of argumentation in terms of persuasiveness has been investigated by several researchers (Carlile et al., n.d.; Lawrence & Reed, 2019) – while *logos* might be the simplest persuasive strategy to program because of its logical base, *ethos*, or the character of the speaker, might be another interesting aspect to consider. *Ethos* is a crucial feature of political debate, as politicians long realised that building a character might be more
useful than simply appealing to the logical reasoning of the policy in question. The large amount of political discourse available online, either in the form of recordings of political debates or opinion pieces in blogs or social media, requires any attempt to model or mine arguments for further analysis to consider *ethos* and *pathos* as much as *logos*. Several works by (Duthie et al., 2016; Duthie & Budzynska, 2018) have tried to bridge this gap by studying *ethos* mining towards computational argumentation applications, by using recursive neural networks (RNN), sentiment analysis and identification of the speakers by recognising anaphors commonly used in the UK’s parliamentary debates (Duthie & Budzynska, 2018). These researchers were interested in classifying the instances of a politician talking about another politician’s or party’s *ethos* and whether these instances represented a positive or negative sentiment, also called ethotic sentiment. They cross-correlated their results, finding correlation between changes in Margaret Thatcher’s *ethos* and relevant events in her political career. These authors also used a knowledge base to find out the identity of politicians only mentioned by their roles (e.g., “Prime Minister”) at specific time points. This paves the way for a more holistic approach towards *ethos* mining that also considers the audience’s perspective, by identifying the ethotic sentiment of the speaker, but also the general *ethos* of the target scraped from knowledge bases, which could be different from one another. This work comes closer to grasping the nuances between the belief system of the speaker and that of the general public. Ethotic sentiment of a sentence classified using such techniques could be an interesting feature allowing better understanding of the relationships between speakers in political debates or other types of argumentation.

In political discourse, emotive rhetorical features can also be highly dependent on the nature of the deliberation and the audience needing to be appealed to. (Osnabrüge et al., 2021) found that the UK’s legislative debates with a large audience strategically increase their emotive rhetoric to appeal to the wider audience. (Slapin & Kirkland, 2020) used statistical techniques (logit, random forest) to correlate features of linguistic complexity like length of sentences or rarity of words, as well as sentiment categories from dictionaries like the Lexicoder (L. Young & Soroka, 2012) and Emolex (Mohammad & Turney, 2013). While the combined effort of using these quantitative techniques to understand political discourse presents a step forward towards real interdisciplinary research between both disciplines, the state of the art in natural language processing and dialectical models of deliberation within computer science goes further than these that currently bear the standard for the best interdisciplinary work on political dialogue. Nevertheless, there have been efforts to model rhetorical argumentation in a way that could be used by automated models in computer science to build better multi-agent systems.

**Dynamic modes of discourse – a final word on storytelling and changing frames of discussion**

Narratives have a great power in shaping human communication, beliefs and identity forming (K. Young & Saver, 2001). Narratives, and in particular story-telling, have been historically employed in political context as a means of public engagement, but their use in the context of political debates within Parliament continues to be under-researched, as per (Prior, 2018). (Jones & McBeth, 2010) introduced a Narrative Policy Framework (NPF) approach to study policy narratives with an empirical basis. In their work, they emphasise the debate between positivist and postpositivist approaches to public policy studies, resulting in narrative policy approaches of structuralist and poststructuralist nature. Structuralists focus primarily on the text as unit of analysis (Jones & McBeth, 2010) and they aim at generalising narrative aspects, while poststructuralists critique these categorisations by claiming that facts, more often than not, are social constructs that cannot be separated from individual interpretations. Irrespective of the approach, narratives and stories have an undeniable importance in policy controversies and political debates and a deliberative discourse can make use of both
argumentative and narrative techniques to present a point. Rhetorical analysis of texts (especially those regarding ethos and pathos) in computer science, namely in MAS and argumentation mining, share similarities with the study of narratives in philosophy of language. It is clear that a structuralist/positivist approach is favoured in computer science applications, but input from historically more qualitative disciplines like large parts of political studies or philosophy provides an interesting way to advance the field. Therefore, as the use of non-standard ways of communicating (story-telling) is understudied in this manner, the use of computer science techniques should be a fantastic tool. Adopting a dynamic view on facts does not imply denying the existence of truth, but simply understanding that construction of knowledge (and its current state at the time of interaction) influence and determine the speaker and the public’s view of reality and, overall, their behaviour and understanding (their pathos, their ethos and even their logos). Advances in digital humanities here can encourage researchers who rely usually on qualitative evaluations of democratic innovations to embrace computational methods. Frame theory, for example, has been used in conjunction with argumentation schemes as a type of classification that can be extrapolated to computer science and classification tasks, for instance entity recognition. According to Wohlrapp, the frame in a discussion can shift, and with that shift the premises might lose their power or might change from objection to support when new information is added. This type of dynamism is not usually taken into account when designing or implementing argumentation frameworks, but it could have a relevant impact. As Wohlrapp puts it: “For these steps we certainly need more than logic: we need an open mind, creativity and mental awareness.” Can computers really capture this level of complexity or richness? Richer models of deliberation have proposed ways to extend McBurney’s model into one that can capture dialogue shifts when new information appears or when closure is necessary, and a conclusion has not been reached (Walton et al., 2016). Although many refinements as this one keep being investigated to establish novel and more inclusive argumentation frameworks, most of them still lack features that others consider vital. Some lack open knowledge bases, while others lack flexibility or dynamism to adapt to changing circumstances, and others lack the use of purely rhetorical figures like pathos or ethos. As of now, every computerised interacting system still needs significant human intervention and can only be used in specified settings. A broader question still remains: will they always need human intervention?

Knowledge representations and algorithms should be rethought in a way to allow a certain dynamism and change of circumstances, relative views of the world or the social construction of reality, perhaps mid-way between the New Rhetoric and argumentation frameworks based on logical microstructure of the arguments. While purely logical frameworks might be useful for interacting agents that are based within the same logical principles, interaction with humans will require modelling and adapting to their discourse, which includes story-telling, narrative, dynamic knowledge and different degrees of reality construction. Collaboration between disciplines in a holistic way, instead of purely applying techniques without extensive result interpretation, could result in a much more rewarding and useful experience for researchers. Validation of algorithms that does not stem only from accuracy metrics, but that considers their ability to describe reality, to adapt and to construct knowledge alongside their users, is crucial for practical applications that go beyond the computer science realm.

A rhetorical framework thus aims to identify discursive techniques that might have a significant effect in a dialogue and to model the argumentation process by focusing on the audience and speaker’s perception, and not on the argument itself. Here, we have extended this setting to consider other features like shifts in the conversational frame or encountering new information. Fully rhetorical frameworks have not been extensively applied, due to the difficulty of finding automated methods to analyse these rhetorical figures. These types of frameworks usually also rely on the
concept of argument strength, which on the other hand can be easier to implement in computer science applications.

5. Conclusion

Our work here was inspired by an observation that several disciplines have been working separately in efforts to improve understanding and practice of political communication through modern data science. There is a clear lack of knowledge (from political scientists and computer scientists in particular) of what is being done in other fields including in the legal domain, education, etc. This is evident in the widely different ontologies employed by both fields, with different naming conventions for what are essentially the same phenomena. This occurs even when there are touchstones that are common to fields – e.g. fundamental or popular theories of communication such as the work of Habermas.

When political science and computer science have worked together to develop tools to assist in deliberation or argumentation, they tend not to develop very user-friendly tools or lack focus on goals that are useful to political actors. This happens partly because they speak different languages and are incentivised by their own disciplinary goals, but also because the real trade-offs between complexity and parsimony in modelling speech and their consequences have not been systematically explored.

In this paper we provided several examples of how far collaborations have come, what theories they draw on provide food for thought identifying the most purposeful future interdisciplinary collaboration that ought to take place. The intersection of communication data and novel communication technologies currently poses a problem for democracy because much political communication happens on platforms built for alternative (and often extractive) purposes. We should not be surprised when these systems undermine democracy. Rather than withdraw from the challenge, we ought to think about how advanced computational power can more properly model political communication such that we can put power in the hands of many more people to use novel technologies to augment their political power. We find that the most important avenue for research is to understand better how computer science could take advantage of the use of ethos, pathos and dynamic modes of discourse. Rather than focus on narrow research questions, we entreat colleagues to develop models that consider each of these aspects of communication more often and help us learn lessons more quickly to keep up with the pace of change.

References


Carlile, W., Gurrapadi, N., Ke, Z., & Ng, V. (n.d.). *Give Me More Feedback: Annotating Argument Persuasiveness and Related Attributes in Student Essays*.


