

Building more democratic institutions with collaboratively constructed debate moderation tools

By Max Holschneider

One of the most important underpinnings of democracy is debate. Debate lets us discuss ideas, examine viewpoints, and, as a collective (after due process), make decisions that we believe to be right. Since the time of the Ancient Greeks, healthy debate has been the foundation of healthy democracy, with transparent rules of engagement that are carefully and sacredly followed. Belief in these systems was what led to the birth of our democratic republic—for example, the House of Burgesses, established in 1619 in colonial Virginia, was one of the first representative assemblies in the Americas. It provided a forum for elected officials to debate issues and make decisions on behalf of their constituents, setting a precedent for the democratic institutions that would later shape the United States.

However, in recent years, we have witnessed a troubling shift in the nature of political debate. While never having been perfect, empirically there has been a marked decrease in the quality of political debate. One need look no further than to our leaders, where events like the Republican and Democratic debates have devolved into shouting matches, false accusations/fake news, and candidates rarely providing substantive answers. This is not without reason; our candidates are simply playing the best they can under the given rules of engagement. These issues are by design—the problem lies in the fact that the rules by which these debates are conducted, which should be designed to foster productive and informative discussions, are often poorly defined or unevenly enforced.

In this paper we will present a framework for a debate moderation system collaboratively constructed by voter input. We will describe the central issues surrounding our current structures of debate, and how such a framework could address said issues. Building out a collaboratively constructed debate moderation system could better represent the will of the people that elected officials are vying for. Lastly, we claim that such is feasible with today's technology.

Issues:

Our current systems of debate inherently optimize towards the wrong directions. As it stands, the incentives of current debate do not put the best interests of the public at its forefront. Instead it is the economic and ideological interest of other organizations that take the driver seat. As a result, not only do we have corrupted intentions, but we get an outcome where we have debates that are poor quality, and do not hash out the kitchen tables issues which Americans are interested in.

So, who defines these rules of engagement? Nowadays, political debates are often led by news outlets. Networks like Fox, CNN, and C-SPAN often select debate moderators from their own teams, who we would like to act as impartial arbiters to enforce predefined rules. However, this rarely happens.

News outlets themselves are highly polarized and often have personal interests. For instance, Rupert Murdoch, the owner of Fox News, has been known to use his media empire to promote his political agenda and influence public opinion. While Rupert Murdoch is an important example due to his outspokenness about using his influence to propagate his ideology,

oftentimes such claims are difficult to prove. What interests are very easy to prove however? Financial ones. Well-structured, honest debate doesn't generate clicks or views. In the current media landscape, news outlets are driven by the need to attract and retain viewers or readers for individual clicks above deep engagement. A debate that adheres to strict rules and focuses on policy issues may be informative, but it just doesn't bring in the ad dollars the way two presidential candidates yelling over each other for an hour does.

Why do we care?

The issues with the current debate structure extend beyond the conflict of interest of news outlets however—the public feels disconnected and disengaged from the debate process itself. The rules are often poorly defined, perhaps paid lip service to at the beginning of a debate, but beyond that, one can only trust the arbiter. In this a large disconnect becomes evident—why should a voter care if a candidate violates a rule that they have no say in creating or enforcing?

We find ourselves in a situation where we, as the people, can't decide the fuels of engagement, can't have a say in the enforcement of said rules because the arbiter has sole control over whether or not they are enforced, and the arbiter is selected by groups with strong conflicts of interest. Is it any wonder that individuals feel disenfranchised from witnessing their future elected leaders debate?

Solution:

Our proposed solution centers around creating collaboratively constructed debate moderation systems. In essence, we argue that a debate moderation system in which the people

decide the rules of debate, and is built to fairly enforce those rules is not some high-minded fantasy, but a feasible goal. Our proposed solution has two major domains of mechanism design:

1. *Building a robust system where we can collectively agree on rules.*

We look to develop voting systems that allow people to evaluate different styles of debate rules, considering the pros and cons of each, before the debate takes place. This would enable the public to vote on which rules they would like to see implemented. These voting systems could be built to integrate with existing social media polling technology, ensuring accessibility for a wide range of citizens. To maintain fairness, the voting process should be transparent, with clear guidelines on who is eligible to vote and how the results will be used to shape the debate rules. Through building a forum where rules can be voted on, we can empower existing institutions such as academia in presenting arguments for better debate design.

2. *Constructing the systems that actually enforce those rules.*

While not exhaustive, here is a preliminary list of a few types of rules and how they could be computationally enforced:

- a. Ensuring speakers do not exceed their allotted time: This is simple to implement. Each speaker has a timer, and when their time is up, their microphone is cut off, and the next candidate gets to answer. One could explore less rigid systems, perhaps allowing candidates to go overtime a limited number of times during the debate. The key is that the rules are clearly defined and determined by the people.

- b. Addressing the issue of candidates not providing specific answers to questions that have clear binary responses, such as “Are you for or against the enactment of bill SB9 in California?” In these cases, a GPT-like system could be used to evaluate candidates’ responses and impartially call out any that do not provide proper answers. For further clarification, we have built a toy example of this in our linked GitHub repository.
- c. Real-time fact-checking: By using data from trustworthy sources like PolitiFact or FactCheck.org, we can build agents that fact-check arguments and add context when something is flagrantly false. These sources have a proven track record of objectively evaluating political statements. For example, during the 2016 U.S. presidential election, PolitiFact found that 70% of Donald Trump’s statements were mostly false, false, or “pants on fire”, while 26% of Hillary Clinton’s statements fell into those categories. Integrating such data into real-time fact-checking systems would help to ensure that candidates are held accountable for their words. Challenges obviously remain, such as the need for rapid access to reliable information and potential disagreements over what constitutes a “fact.”

Systems like these won’t be perfect from the outset. The scope of this paper is not to provide the exact techniques that will create the perfect debate structure but rather to argue that the collaborative construction of debate moderating agents, which can serve as better representations of the people’s will compared to private interests, is not an impossibility. This paper does not tackle the questions of partiality, and it is important to note that these systems will be inherently vulnerable due to the partiality of an inherently political system. Further, there is

clearly an important interplay between building systems that can act as debate moderators and ensuring that we build those systems correctly. These systems will make mistakes, and we may vote for debate rules that create similarly bad conditions to the political debates we see today.

But these issues are addressable—and a transparent, iterative system could significantly enable individuals to participate in building more robust debate systems. As the public engages with the debate process and observes the performance of the moderating agents, they can provide input on areas for improvement, in an open-source manner. This feedback can be used to help us build and use better rules.

If we can build debate systems in which the rules of moderation are decided by the people, people are far more likely to care when those rules are broken. Every election season we hear that the debates are rigged, so why not build them ourselves?