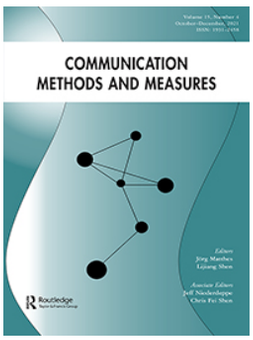

Communication Methods and Measures

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hcms20

Good Decisions or Bad Outcomes? A Model for Group Deliberation on Value-Laden Topics

Sarah Shugars

To cite this article: Sarah Shugars (2021) Good Decisions or Bad Outcomes? A Model for Group Deliberation on Value-Laden Topics, Communication Methods and Measures, 15:4, 273-291, DOI: 10.1080/19312458.2020.1768521

To link to this article: https://doi.org/10.1080/19312458.2020.1768521

COMMUNICATION METHODS AND MEASURES
2021, VOL. 15, NO. 4, 273–291
https://doi.org/10.1080/19312458.2020.1768521

Routledge
Taylor & Francis Group

Good Decisions or Bad Outcomes? A Model for Group Deliberation on Value-Laden Topics

Sarah Shugars

Network Science Institute, Northeastern University, Boston, Massachusetts, USA

ABSTRACT

Agent-based models present an ideal tool for interrogating the dynamics of communication and exchange. Such models allow individual aspects of human interaction to be isolated and controlled in a way that sheds new insight into complex behavioral phenomena. This approach is particularly valuable in settings beset by confounding factors and mixed empirical evidence. The political communication setting of deliberation is one such salient setting: in business, politics, and everyday life, individuals with varying opinions, experience, and information attempt to collaborate and make decisions. Empirical evidence suggests that such collaborative reasoning can lead to good decisions, yet there are numerous deliberative failures which may frequently cause groups to reach bad outcomes. Using the substantive setting of deliberation, this paper presents an agent-based model aimed at disambiguating the individual factors which influence decision-making conversations. We model this communicative process as a deliberative game of "giving and asking for reasons." Agents share beliefs around possible policy initiatives and attempt to enact "good" policies through a process of mutual exchange and consideration. The model considers an interconnected policy landscape in which implementing or not implementing a policy mediates the value of other policies. Within this framework, the paper considers the impacts of three canonical failures of deliberation: limited cognitive capacity, group factions, and tendencies to make poor judgments when accepting or rejecting others' views. We find that cognitive capacity can significantly decrease the ability of a group to reach a good decision. However, this effect appears to be mitigated for groups of opposing factions. Indeed, polarized groups do surprisingly well at identifying optimal policy solutions, suggesting that heterogeneous agents can achieve good outcomes *if* they are willing to talk and learn from each other.

Introduction

Collaborative reasoning – whether successful or unsuccessful – is a core facet of human society. In business, politics, and everyday life, individuals with varying opinions, experience, and information attempt to collaborate and make decisions. However, the ability of these groups to come to "good" decisions may be hindered by both group dynamics and individual failings. Such challenges are a core concern among detractors of deliberative democracy who argue that it is overly idealistic to imagine everyday citizens capable of successfully reasoning together about matters of common concern. Such efforts are arguably doomed to fail due to individuals' limited cognitive capacity (Lippmann, 1922; Oswald & Grosjean, 2004; Shah & Oppenheimer, 2008; Tversky & Kahneman, 1981), the inability of divided factions to agree (Dworkin, 2006; Madison, 1787), and tendencies for

CONTACT Sarah Shugars shugars.s@northeastern.edu Network Science Institute, Northeastern University, 177 Huntington Ave, Boston, MA 02115-5005

individuals to accept or reject information on inaccurate grounds (Festinger, 1954; Janis, 1972; Nickerson, 1998; Sunstein, 2002).

These concerns about group decision making are heightened in political settings where citizens may hold subjective, value-laden perspectives, and where the "truth" is hard or impossible to know. Such settings are importantly different from deliberation under factual circumstances where individuals may be limited in their knowledge but would fully agree if given access to the same information. In these factual settings, collaborative reasoning can be well-modeled under an explore/exploit framework where agents attempt to find a global optimum given local information (Lazer & Friedman, 2007; Mason & Watts, 2012). However, if the topic is value-laden, if agents hold their own subjective opinions of the solution space, existing models can neither explain nor predict the process of collaborative reasoning.

Yet, a great deal of real-world problems rely upon agents coming to normative judgments. In the political realm, for example, a policy solution is only optimal if it results in outcomes an agent would qualify as "good." Political polarization, in this sense, does not necessarily represent an inability to discover and share relevant information, but rather a deeper disagreement as to the *value* of that information. Despite skepticism to the contrary, numerous empirical studies demonstrate that people are able to productively discuss value-laden matters (Fishkin, 2014; Knobloch et al., 2013; Neblo et al., 2010, 2018), suggesting a growing need to understand the conditions under which these conversations succeed.

Agent-based models present an ideal tool for such a task. First, these models allow us to isolate and interrogate different aspects of individual behavior. While agent-based models cannot capture the full complexities of human behavior, this limitation is also their strength: by intentionally selecting and defining available agent actions, ABMs allow researchers to disregard possible confounders and complex interactions. Such models are, therefore, invaluable in the face of conflicting empirical evidence and hold a unique potential to clarify the impact of factors that could not otherwise be disambiguated. Furthermore, using an agent-based model in this setting allows us to appropriately benchmark deliberative success. Indeed, the ability to accurately measure "success" has been a significant challenge for empirical studies of deliberation (Knobloch et al., 2013; Mansbridge, 2015), where neither subjects nor researchers can be sure whether or not a group has come to the best decision.

For both those reasons, this paper presents an agent-based model of collaborative reasoning on value-laden topics, drawing upon literature on group problem-solving (Lazer & Friedman, 2007; March, 1991; Mason & Watts, 2012) and belief convergence (DeGroot, 1974; Friedkin et al., 2016). In a deliberative game of "giving and asking for reasons" (Neblo, 2015), agents share beliefs around possible policy initiatives and attempt to enact "good" policies through a process of mutual exchange and consideration. The policy landscape is taken to be a complex system in which the implementation or non-implementation of individual policies has non-linear effects on the value of an overall policy platform. In order to assess the ground truth "value" of deliberative outcomes, this model considers settings in which members agree as to the best overall outcome but disagree as to the specific policies which come closest to achieving that outcome. For example, we might imagine citizens who all want to live in healthy, safe, communities with access to good education, but might also imagine those same citizens disagreeing on both the degree to which those goals are obtainable and the set of policies which come closest to bringing about those goals. In such a setting, the ground truth solution would describe the optimum set of policies needed to achieve the best possible outcome.

The model, described in detail in Section 3, considers a solution space or "policy landscape" in which the implementation or non-implementation of a policy affects the value of other policies. Specifically, the model leverages the *NK* solution space in order to describe a set of N policies whose values are each mediated by K other policies. Initially conceived of in the context of adaptive evolution (S. A. Kauffman & Weinberger, 1989; S. Kauffman & Levin, 1987), *NK* models have been widely applied to organizational and group problem-solving tasks (Geisendorf, 2010; Herrmann et al., 2014; Lazer & Friedman, 2007;

Levinthal, 1997; Shore et al., 2015). Here, we extend this model to an important communications challenge, examining political communication and deliberation. In doing so, we introduce a novel methodological framework for examining the exchange of messages and the adoption of beliefs.

Using an agent-based model for this task allows us to isolate and examine several core concerns about the practicality of deliberation. Specifically, this paper considers deliberative outcomes for imperfect agents who are prone to a number of individual and group failures. The mere existence of such failures are often taken to imply that deliberative success is an unrealistic and unachievable goal (Dworkin, 2006; Lippmann, 1922; Sunstein, 2002). However, the extent to which any one of these failures dooms deliberation or fits within a broader deliberative system (Mansbridge, 1999) has remained largely unaddressed. Specifically, this paper aims to determine the deliberative impacts of three canonical failures of deliberation: limited cognitive capacity, group factions, and tendencies to make poor judgments when accepting or rejecting other's views. As described in Section 3, uninformed agents reflect concerns that individuals have limited cognitive capacity and are therefore unlikely to hold views closely reflective of any underlying ground truth (Lippmann, 1922; Oswald & Grosjean, 2004; Shah & Oppenheimer, 2008; Tversky & Kahneman, 1981). Ideologues reflect concerns that polarized factions will remain entrenched and divided with conflicting viewpoints (Dworkin, 2006; Madison, 1787). Finally, a tunable open-mindedness parameter captures both the effects of groupthink and confirmation bias, as individuals are either too quick to accept group views (Festinger, 1954; Janis, 1972; Sunstein, 2002) or, alternatively, reject views which do not closely conform to their existing priors (Nickerson, 1998). This paper, of course, does not address all concerns about deliberation. Most notably, this model assumes that all agents participate openly and equally and therefore does not consider that people may generally shy away from conflict (Eliasoph, 1998; Mutz, 2006), may have little interested in engaging (Hibbing & Theiss-Morse, 2002), or may be shut out of deliberative discussions by those with more power (Gaventa, 1982; Sanders, 1997).

Running this model with small groups of deliberators, we find that ideologues – whose skewed beliefs reflect concerns about factions and polarization – do surprising well at identifying optimal policy solutions. Indeed, while concerns about individuals' cognitive capacity appear to be well-founded, factions of oppositely-skewed peers can overcome this limitation in settings where they are willing to listen to each other. This suggests that factions – while ostensibly contentious – may actually balance each other out and come to an optimal middle ground.

Related Work

A long line of work has leveraged agent-based models to examine the dynamics of group consensus and dissensus. Capable of interrogating and isolating the myriad elements which influence group decision making, these stylized models have proven to be a fruitful supplement to human-subject experiments. This work was largely pioneered by DeGroot (1974), who imagined small teams tasked with reaching consensus about the ground truth value of some parameter, θ. In this model, each agent holds a unique probability distribution as to the value of θ and also assigns a value to the opinion of each other agent. Trust is a major variable of interest in this model, and, indeed, DeGroot (1974) finds that if a single agent's opinion is positively valued by all players, the group's opinion will converge.

March (1991) more explicitly examines the organizational context, presenting a model in which individuals are socialized to an organization's beliefs while the organization simultaneously learns from its members. March (1991) frames this as an explore/exploit trade-off where individuals and the organizational code may either converge quickly and exploit suboptimal solutions, or converge slowly and explore better solutions. March finds that the inefficiency introduced by the presence of "slow" learners helps the system converge optimally rather than settling on a local peak. This finding is echoed in more recent work on solution convergence in group problem-solving tasks. In work that also uses the *NK* model as a solution space, Lazer and Friedman (2007) find a notable "trade-off

between maintaining the diversity necessary for obtaining high performance in a system and the rapid dissemination of high-performance strategies." While the human-subject experiments of Mason and Watts (2012) provide partial validation for the results of Lazer and Friedman (2007), they notably find that that human subjects were able to simultaneously benefit from a diversity of opinions and rapid dissemination. This suggests that while rapid dissemination allowed subjects to adopt others' approaches, users were set enough in their ways that this mechanism didn't automatically result in convergence on sub-optimal solutions as simulations predicted.

Models of consensus and belief systems are particularly relevant for deliberative theory, where decision making is often taken to be a required outcome (Mansbridge, 2015) of an exchange of reasons (Gutmann & Thompson, 1998; Mansbridge, 2015; Neblo, 2015). In this sense, political deliberation can be taken as a group problem-solving task, where participants come in with private information and search for the optimum of a complex solution space through an iterative process of search and knowledge exchange. In this task, deliberative agents face a similar explore/exploit tradeoff: good faith discussants ought to "convince others and be convinced when appropriate" (Mercier & Landemore, 2012). That is, a person engaging in political discourse ought to aim to believe *true* things – but doing so requires a careful balance between intellectual humility and self-confidence. When encountering another person's viewpoint, a good-faith deliberator ought to carefully consider that view, assess its accuracy, and then either adopt this alternative view or provide a reasoned explanation as to why the interlocutor is wrong. While this may not be the most common form of political discourse, it is worth noting that such reasoned exchange can occur (Fishkin, 2014; Knobloch et al., 2013; Neblo et al., 2010). Indeed, such productively collaborative exchange is the backbone of academic discourse.

Explicitly bringing belief systems to the deliberative domain, Altafini (2013) considers the case of "agreed upon dissensus" or "bipartite consensus" where antagonistic agents converge on the same values but with different signs. In political talk, this constitutes the case where deliberators "agree to disagree" – each finding the other's argument to be reasonable, rational, and wrong. Choi and Robertson (2013) further model belief systems in collaborative governance. Comparing the outcomes of deliberation across various decision rules (unanimity, supermajority, and dominate coalition), they find that deliberation is more important than voting rules in building consensus and enhancing decision quality. Friedkin et al. (2016) study belief systems with logic constraints; examining the convergence of opinion when arguments are rationally linked. They find that a strongly linked logic structure can shift people away from their initial beliefs, though the presence of unrelenting critics can mitigate this effect.

Developing an Agent-Based Model for Deliberative Exchange

This paper imagines small groups of citizens deliberating about a set of interconnected policy solutions that all influence an overarching social issue. Agents are assumed to agree as to the ideal overall outcome, but disagree about the specific policies which best move toward that ideal. For example, a group discussing public education would share a desire for students to get a good education but might disagree on what makes an education "good" or on which specific policies meet the needs of the most students. A group discussing healthcare would agree that, ideally, everyone would have access to affordable healthcare, but they would disagree as to feasible solutions for providing the best healthcare to the most people. Similarly, a group discussing crime might all hope to see a reduction in crime while also disagreeing as to what policies are most likely to bring about such a reduction.

While agents share an overarching goal and consider a common set of possible policies, they hold differing views as to the value and implication of each policy under consideration. These individual-level views, which may be driven by normative or practical considerations, lead agents to begin deliberation with different beliefs about the best set of policies to implement. During deliberation, agents take turns sharing and considering reasons for why a policy should or should not be

implemented. These reasons, which will be described in detail in Section 3.3, can be intuitively understood as the broader implications of implementing a policy, beyond the value of that policy itself. After each exchange, agents give every policy an individual up or down vote, resulting in an implemented policy platform. The value of the implemented platform is then judged against a ground truth solution.

This model, then, requires three basic components which will be described in detail below: (1) a ground truth solution space against which we can benchmark deliberative performance, (2) individually-initiated belief spaces representing the views of each deliberating agent, and (3) a process of reasoned exchange in which agents share views and decide whether to accept the views of others. Within this framework, this paper evaluates the impact of three canonical deliberative failures. Concerns about the limits of humanity's cognitive capacity (Lippmann, 1922; Tversky & Kahneman, 1981) are tested by manipulating the initial accuracy of individuals' beliefs. Fears about the inability of polarized groups to productively collaborate (Dworkin, 2006) are examined by inducing coalitions which are divided in their initial beliefs. Finally, the implications of deliberators being too easily swayed by peer information (Janis, 1972; Sunstein, 2002) or too unwilling to consider alternative views (Nickerson, 1998) is determined by tuning a parameter governing agents' acceptance or non-acceptance of information. These parameters and experiments are described in detail in the following sections.

Ground Truth Solution Space

A core assumption of this model is that policies can't be fully understood in isolation; that they are deeply interconnected. In other words, the value of any single policy is mediated by whether or not *other* policies are implemented. For example, in public education, a group might consider policies targeting student testing, teacher evaluation, or free and reduced lunch programs. However, the success of any of these policies may depend on the implementation – or non-implementation – of the other policies under consideration. The reliability of student testing may be influenced by the existence of a free and reduced lunch program while the value of teacher evaluations may be influenced by practices for student testing. Within the healthcare domain, a group may consider a mandate for health insurance separately from the implementation of a single-payer system, but the success of either of these policies may rely on the implementation of the other. Similarly, a group aiming to reduce crime will not only have to grapple with questions of police officer training, patrol strategies, and incarceration policies, but also consider how those policies interact with each other. For example, increasing patrols may require additional training in order to be effective.

In order to capture the interconnected implications of policy implementation, the model leverages the *NK* solution space for both the ground-truth solution and individuals' beliefs. Initially developed for adaptive evolution (S. A. Kauffman & Weinberger, 1989) and widely adopted for group problem solving (Geisendorf, 2010; Herrmann et al., 2014; Lazer & Friedman, 2007; Levinthal, 1997; Shore et al., 2015), the *NK* model allows us to capture both positive and negative influences between policies. Each of N policies can exist in one of two states: either implemented (state = 1) or not implemented (state = 0). A system with $N = 4$ policies under consideration would, therefore, have $2^N = 16$ possible policy platforms – unique combinations of policies solutions – ranging from 0000 (no policies implemented) to 1111 (all policies implemented). The complexity of this landscape is then moderated by the parameter K, which indicates the number of policies whose state influences the value of implementing or not implementing a considered policy. This means that the overall value of the policy platform 1000 differs from the value of policy platform 1001 not only by the "raw" value of the 4th position policy but also by the influence that policy has on the value of other policies.

Figure 1 shows example influence patterns for an *NK* model in which $N = 4$ and $K = 2$. The contributing value of any given policy is determined by $K = 2$ other policies, but this relationship is unidirectional – e.g., as shown in the Figure 1 example, the contributing value of policy D is influenced by

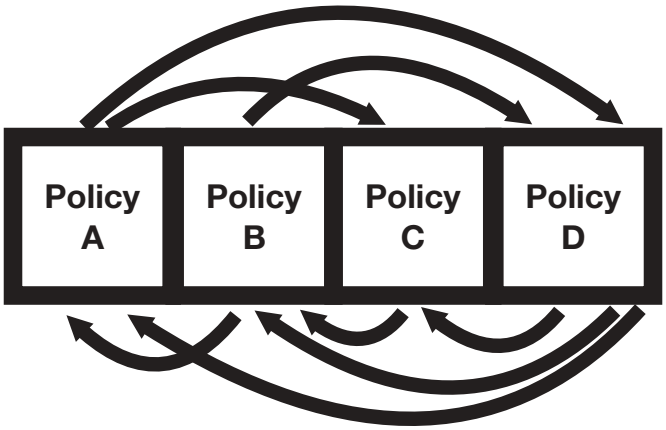

Figure 1. Example influence patterns N = 4, K = 2.

policies A and B, while the state of policy D (whether it is implemented or not implemented) influences the contributing value of policies A, B, and C. Thus, some policies may have an outsized effect on the overall policy platform, while other policies may exert little influence.

Given these influence patterns, the contributing value of a single policy to the overall policy platform is then determined by the state of that policy (e.g., whether it is implemented or nor implemented) as well as the states of the *K* influencing policies. Table 1 shows example influence

Table 1. Example influence values for *N* = 4 policies.

Policy	State	Influencer	Influencer State	Influence Value
A	0	B	0	3.39
			1	−3.16
		D	0	−4.97
			1	3.08
	1	B	0	8.87
			1	5.28
		D	0	7.56
			1	7.22
B	0	C	0	−9.69
			1	−5.86
		D	0	7.9
			1	−1.03
		C	0	−1.0
			1	−9.75
		D	0	5.83
			1	−5.6
C	0	A	0	−0.39
			1	−4.1
		D	0	2.77
			1	9.22
	1	A	0	8.79
			1	−5.22
		D	0	6.64
			1	0.52
D	0	A	0	−5.6
			1	2.09
		B	0	7.48
			1	8.03
		A	0	3.63
			1	−5.4
		B	0	−7.71
			1	5.14

values, given the influence patterns described in Figure 1. For example, the contributing value of policy A is determined by the states of policies B and D. Thus, if policy A is implemented (state = 1) and policy B is not implemented (state = 0), the B → A relation has a value of 8.87. Similarly, if policy A is implemented (state = 1) and policy D is also implemented (state = 1), the D → A relation has a value of 7.22. Averaged together, this means that the total value of having policy A implemented in this policy platform is 8.045. Note that in this example, whether or not policy C is implemented will have no effect on the contributing value of policy A.

The results presented in this paper take the strength and valence of each policy interaction as a random draw from the uniform distribution unif(−10,10) though these results are robust across arbitrary bounds on this distribution. Negative values are assumed to be detrimental to the overall goal (e.g., weaken public education), while positive values are taken to indicate a policy is supportive of that goal (e.g., strengthen public education).

Taken together, these interactions assign a value to each of the 2^N possible policy platforms. The contributing value of any given policy is determined by its own state (whether it is implemented or not implemented) as well as the state of K other policies. The total value of a given policy platform can then be calculated across all contributing policies. An example of this calculation is shown in Figure 2.

Using the *NK* framework, the model begins by initiating ground truth parameters of influence patterns (which policies influence each other, as in the Figure 1 example) and influence values (the strength and valence of that influence, as in the Table 1 example). Together, these influence patterns and values provide ground truth measures for every possible combination of policies. For all simulations presented in this paper, we take $N = 4$ and $K = 2$. This is a relatively simple terrain which captures the inherent complexity of policy decisions while allowing for a sharper focus on the conditions under which political discussion may be productive.

Conceptually, the ground truth here reflects the true set of policies that best achieve the overarching social issue being discussed – i.e., the set of policies that best support public education or do the most to reduce crime. Note that the need to benchmark deliberative success against a ground

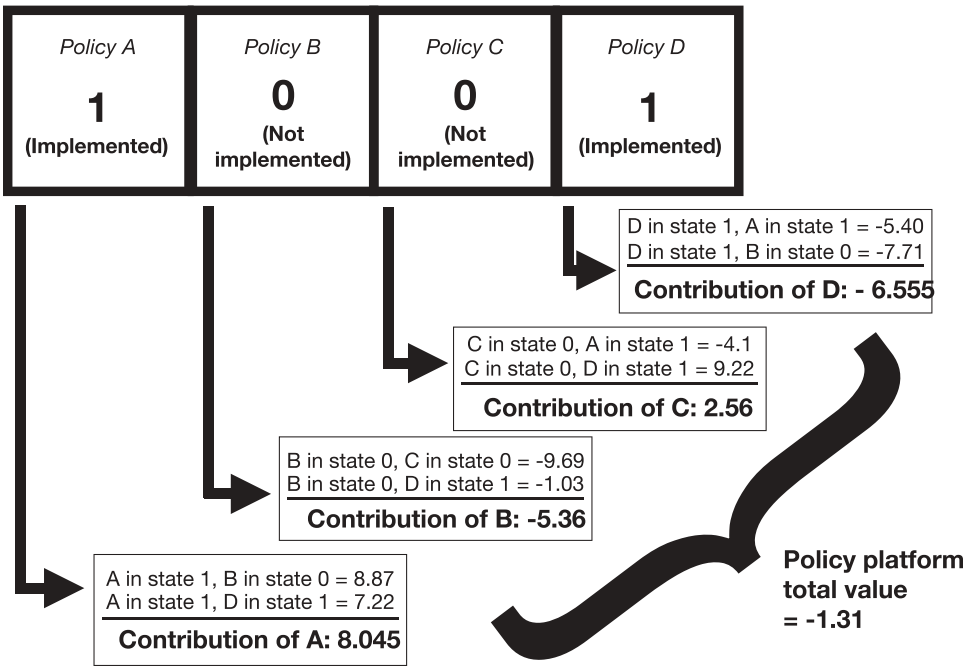

Figure 2. Example policy plateform calculation.

truth solution limits the model to only considering topics where agents broadly agree on the ideal outcome. That is, agents are assumed to largely want the same things – for public education to be good or for crime to be reduced – but to have differing normative and practical views as to which individual policies are best poised to achieve that shared goal. Agents themselves do not have direct access to this ground truth: such a solution may exist but be difficult to know or may be best considered as the outcome of some prior social process.

Indeed, this is one of the great challenges of assessing the quality of real-world political conversation: neither researchers nor deliberators typically have ground-truth knowledge as to what solutions truly are best. One of the benefits of agent-based models, then, is that we can arbitrarily define the "ground truth" and define agents' behavior in relation to that truth. This approach assumes that there is some meaningful benchmark against which to compare agent performance, but it is agnostic as to the social or natural processes which determined that benchmark.

Individual Beliefs

Individual agents are assumed to share an overarching goal, but begin deliberation with differing normative and practical beliefs about the best policies for achieving that goal. In the language of the *NK* model, all agents consider the same *N* policies and share the same set of ground truth interaction patterns (as in Figure 1). However, agents perceive different values for these policy interactions (as in Table 1) and thus favor different policy platforms.

These individual-level beliefs can be understood as noisy interpretations of ground truth signals. That is, while agents do not have direct access to the true, underlying influence values between policies, agents' initial beliefs are assumed to have been shaped by information – such as scholarly research, policy briefs, or direct experience – which may contain a signal of the ground truth value. The closeness of an agent's beliefs to the ground truth can, therefore, be interpreted as reflecting, to some degree, an agent's cognitive capacity – e.g., their ability to synthesize an accurate view from the cacophony of information around them. While operationally, agent beliefs are generated by adding noise to the ground truth influence values, as described below, the assumption here is that agents who are good at processing complex information will hold more accurate (less noisy) beliefs, while those with limited cognitive capacity will have less accurate (more noisy) beliefs.

It is worth noting that the model overall assumes some baseline cognitive capacity. That is, all agents are assumed to hold or *believe* they hold reasoned views. Whether an agent's initialized beliefs are indeed reasoned judgments or simply post-hoc rationalizations, the model assumes that agents will defend their beliefs on rational grounds when participating in deliberation. This approach assumes that there are some settings in which people are capable of engaging in reasoned thinking (Evans, 2003; Kahneman, 2011) and aims to better understand the dynamics which lead that reasoning to be successful or unsuccessful. This is not to suggest that well-reasoned discourse is the modal form of political engagement, merely that there is reason to believe such discourse can and does occur (Fishkin, 2014; Knobloch et al., 2013; Neblo et al., 2010, 2018) and it is, therefore, valuable to understand the dynamics of these settings.

If we were to take the more pessimistic view that policy preferences were never motivated by justifiable reasons, we could perhaps focus instead on agents' preferred policy platforms, using social contagion or similar models to determine which platforms gain widespread popularity. While this may be a fruitful line of additional research, the model here is intended to interrogate different deliberative concerns. Building off Lippmann (user1922) and others, the model can best be understood as examining whether or not deliberation can be successful if citizens are too busy, distracted, or confused to hold well-informed opinions.

Specifically, this paper examines the deliberative impacts of three canonical failures of deliberation: limited cognitive capacity, factions, and poor reasoning regarding the acceptance or rejection of beliefs. The effect of cognitive capacity is examined through a noise parameter which is swept during

a full simulation. Additionally, uninformed and informed agents, described below, capture the extremes of reasoning behavior and provide expected bounds for other types of reasoning agents. The effect of factions, e.g., coalitions with differing views, are examined through ideologue agents who begin deliberation with a cohort of ideological peers. The initialization of these agents is described in more detail below. Finally, the tendency for individuals to "follow the group" or to acquiesce to others' views (Janis, 1972; Sunstein, 2009) or, alternatively, to remain closed to views which don't confirm their existing beliefs (Nickerson, 1998) is examined through the use of an "open-mindedness" parameter described in Section 3.3.

Uninformed Agents

Uninformed agents provided a baseline condition for examining the effects of agents' cognitive capacity. Such agents represent concerns that citizens are too inundated by information or too distracted by personal affairs to hold well-informed views (Lippmann, 1922). That is, if agents hold beliefs that are poor reflections of the true world around them, can they come to good decisions by sharing information?

Operationally, uninformed agents are initiated with a noisy version of the ground truth influence values (e.g., as described in Table 1). These are calculated as the ground-truth interaction plus a random draw from a uniform distribution determined by the noise level, a parameter which is swept over the full simulation. That is, for ground truth value v and noise level n, an uninformed agent will hold an influence value of $v + \text{unif}(-n,n)$. Uninformed agents are therefore equally likely to over- or under-estimate ground truth interactions and are considered to be free of ideological skew. This mechanism captures reasoning capacity solely as a function of noise level. As the simulation sweeps the noise parameter, agents go from holding beliefs that are good reflections of the ground truth to holding beliefs that are effectively random. In other words, in low noise systems agents are assumed to have the cognitive capacity to accurately interpret signals relating to the ground truth values. In high noise systems, on the other hand, agents are assumed to be unable to form good judgments based on any signals they may be receiving.

Note that the model assumes homogeneity across the quality of all agents' reasoning – e.g., all agents are equally good or equally bad at estimating ground-truth values. This also implies an absence of any pre-deliberation context; all possible beliefs are equally likely, and there are no inherent factions or coalitions.

Intellectuals

Intellectual agents are the polar extreme of uninformed agents and provide an alternative bounding condition. While uninformed agents become less accurate as noise in the system increases, intellectual agents remain closely bounded around ground truth values, independent of noise-level. These agents represent an unrealistic ideal which is reflective of concerns about the practicality of good deliberation. That is, these agents capture the argument that deliberation is too idealistic and can only work if all agents are able to develop accurate and well-informed views (Lippmann, 1922). These intellectual agents, therefore, provide both an estimation of ideal behavior and allow us to examine whether technocrats or other highly educated elites, can, or are need to, successfully moderate the inaccurate beliefs of other citizens.

Operationally, as the simulation sweeps the noise parameter, an intellectual's influence value will remain $v + \text{unif}(-1, 1)$, for ground truth value v. Thus, these agents can be considered to have high cognitive capacity and to always hold well-informed views.

Ideologues

Finally, ideologues are initialized as cohorts with intentionally skewed views. These views are equally likely to be any given absolute distance from the ground truth, but are biased in terms of the direction of that distance, e.g., either positively or negatively skewed. The result is that ideologues find themselves in polarized systems, entering conversation with a natural cohort of peers who hold

similar beliefs while simultaneously facing an opposing coalition of peers with dissimilar beliefs. Such agents represent fears that divided groups will remain entrenched in their respective positions and will, therefore, be unable to reach good decisions (Dworkin, 2006; Madison, 1787). Note that this approach still assumes homogeneity across reasoning capacity, as all ideologue agents are initialized with the same absolute skew. In other words, all ideologues are wrong; the resulting coalitions are simply wrong in different ways.

Operationally, for ground truth value v and noise level n, ideologues will be initialized as either $(v + (\text{unif}((n-2),\ n) \times \text{sign}(v)))$ or as $(v + (\text{unif}(-n, -(n-2)) \times \text{sign}(v)))$. Including the sign of v in this calculation allows for non-parametric interactions that appropriately complicate agents' selection of optimal policy platforms. Without this term, these agents could be interpreted simply as optimists (who think all interactions are good) or pessimists (who think all interactions are bad). A consensus solution in such a system would then most frequently reflect pessimists settling for the "least bad" solution rather than advocating for a policy platform they genuinely interpret as good.

Reasoned Exchange

The final element of the model is a process through which agents share beliefs and assess whether or not to incorporate the beliefs of others. In this game of "giving and asking for reasons" (Neblo, 2015), good faith discussants should seek to "convince others and be convinced when appropriate" (Mercier & Landemore, 2012) – that is, they should incorporate others' beliefs when those beliefs seem credible and reject others' beliefs when they do not. However, people are highly prone to social bias (Festinger, 1954), and frequently go along with perceived popular views (Janis, 1972; Sunstein, 2002). Furthermore, people are also prone to confirmation bias (Nickerson, 1998) and, rather than go along with the group, may only accept views that already conform to their existing beliefs. While the deliberative ideal imagines participants aiming to carefully consider all viewpoints (Manin, 2005; Mercier & Landemore, 2012), it is unclear whether or not deliberation can succeed given these common biases. In order to examine this final deliberative failing, then, the model estimates the effect of these biases through an open-mindedness parameter which governs whether an agent accepts or rejects a view.

When sharing an opinion, an agent is more specifically sharing their personal influence values for a single policy in a single state. That is, a speaker can be interpreted as sharing a policy preference (the state of a single policy) and justifying that preference based on the influence of other policies under consideration. For example, a speaker might argue that a policy of standardized student testing should only be implemented (state = 1) if a free or reduced lunch program is also implemented. E.g., an implemented lunch program would have a positive effect on the value of student testing, while not implementing a lunch program would have a negative effect on the value of implementing student testing.

Since the value of each policy is determined by the state of K other policies, this means that each agent associates 2^K values with a given policy in a given state. Furthermore, because we assume that all agents share the same set of underlying influence patterns, this provides listening agents a direct point of comparison. That is, a speaking agent and a listening agent both agree on which policies are related to the policy under discussion, but differ in their evaluation of those connections. The speaking agent can, therefore, share their specific valuation of those connections while the listening agent can compare to their own evaluation of those relations. Using the example influence values from Table 1, a speaker might share beliefs about the value of not implementing policy A (state = 0) as the vector [3.39,−3.16,−4.97,3.08], which any listening agent would understand as proposed influence values for [(B in state 0), (B in state 1), (D in state 0), (D in state 1)] respectively.

In other words, this provides a 2^K dimensional space within which agents can assess each other's views. Agents can then choose to accept or reject a shared view within the context of their own beliefs. If an agent is too open to accepting shared beliefs, they are more likely to accept poor

information and ultimately degrade the decision reached by the group (Janis, 1972; Sunstein, 2002). On the other hand, if agents prefer not to accept shared views, they fall into the trap of confirmation bias (Nickerson, 1998) and may fail to abandon their own false beliefs.

Operationally, at every time step, a single speaking agent shares their length-2^K vector of influence values for a given policy in a given state. Listening agents then compare that to their own length 2^K vector and move toward the speaker's values if the cosine similarity between the two vectors is within a range determined by agents' tunable open-mindedness parameter, described below. If an agent "moves toward" a shared view, they adopt a new belief vector which lies halfway between their old view and the shared view. If they do not accept a shared view, they maintain their original belief vector.

Specifically, this paper considers three types of agents: (1) *open* agents who are prone to the failure of group polarization and groupthink (Janis, 1972; Sunstein, 2002) and who are very likely to incorporate beliefs, (2) *skeptics* who are prone to the failure of confirmation bias (Nickerson, 1998) and only accept views which already closely conform to their own, and (3) *moderates* who aim to accept true beliefs and abandon false beliefs (Mercier & Landemore, 2012). Operationally, since cosine similarity ranges from −1 to 1, *open* agents move toward a shared belief if the cosine similarity is greater than −.75. *Moderate* agents move toward an opinion if the cosine similarity is greater than 0, and finally, *skeptical* agents will only move toward beliefs if the cosine similarity is greater than +.75.

Experiments and Measures

This paper examines the effects of cognitive capacity, factions, and acceptance of beliefs on deliberative outcomes. Particular focus is given to the effect of human cognitive capacity, which is one of the most serious concerns raised about the practicality of deliberative success (Lippmann, 1922; Oswald & Grosjean, 2004; Shah & Oppenheimer, 2008; Tversky & Kahneman, 1981). This capacity is broadly captured in the "noise level" of initialized beliefs, with low-noise systems indicating high cognitive capacity and high-noise systems indicating low cognitive capacity. This noise parameter is swept from 0 to 200 over the course of a simulation while ground truth values are always initialized between −10 and 10. Thus as we tune this parameter, we move from agents whose high cognitive capacity gives them beliefs reasonably close to the ground truth to agents whose low cognitive capacity results in beliefs that are essentially random. Agents are initialized to be either uninformed, intellectual, or ideologue, following the rules above. All agents experience the same noise level, although intellectuals, who are illustrative of individuals with high cognitive capacity, remain unaffected by the noise level of the system.

This paper tests the model particularly for small group discussions, i.e., groups between 5– 25 agents and find similar results across group size. Each simulation is run 100 times at each noise level, with the ground truth initiated only once across all runs at a given noise value. Since this creates some spurious oscillation in deliberative outcomes by noise level, we examine the effects of noise over a rolling window of width 10.

For each simulation, a ground truth is determined, and agents' beliefs are initialized using the rules for uninformed agents, intellectuals, or ideologues as described above. This paper examines a number of possible group compositions. Groups composed entirely of uninformed agents or intellectual agents provide baselines against which to compare behavior. The uninformed group, in particular, captures the effect of diminishing cognitive capacity as the noise level increases. In order to capture faction dynamics, groups composed entirely of ideologues are evenly split between agents with positive and negative skew. In groups with an odd number of agents, the positive-skew coalition is arbitrarily a single agent larger. Additionally, to examine the possible moderating effects of intellectual agents, we consider groups that have a majority of intellectuals as well as groups that have a majority of ideologues. In majority intellectual groups, half of agents are intellectuals, rounding up in systems with an odd number of agents. The remaining agents are split evenly between ideologues with positive and negative skew. In majority ideologue groups, only a single agent (odd systems) or two agents (even systems) are intellectuals while the majority of agents are

split evenly between ideologues with positive and negative skew. The model can further be tested with arbitrary mixes of ideologue and intellectual agents, but the results presented here are illustrative of any group composition.

Once initialized, at each timestep, a randomly selected speaker shares a vector of influence values for a random policy in a randomly selected state. This can be considered a single speech act, with the speaker making a claim about the benefit of implementing a given policy, and then justifying that claim with the influence values which inform that view. Listeners then compare these stated reasons with their existing beliefs, choosing to either move toward the shared values or maintain their existing vector of values, based on the open-mindedness of the simulation. For simplicity, this paper only considers simulations in which all agents have the same open-mindedness rules for incorporating a speaker's beliefs. However, the paper examines three distinct levels of open-mindedness. Open agents represent concerns around groupthink and group polarization (Festinger, 1954; Janis, 1972; Sunstein, 2002), and are likely to accept other's views. Skeptical agents capture issues of confirmation bias (Nickerson, 1998) and only accept views that are already close to their own. Finally, moderate agents fall in between and aim to accept true beliefs and reject false beliefs (Mercier & Landemore, 2012). The paper considers this open-minded parameter particularly for groups of ideologues.

After each speech act (e.g., timestep), the model then assesses the outcome of deliberation by determining how many "good" policies would be implemented if agents held a simultaneous vote on each policy. Each agent selects the policy platform they expect to have the best outcome, given the positive and negative trade-offs the agent sees between different policies. Each individual policy is then simultaneously given an up or down vote, with agents voting for or against the policies indicated by their preferred platform. A policy that receives a simple majority of votes is then considered implemented, resulting in a platform of enacted policies. Note that each policy is given an independent vote with only two options (implement or do not implement) and, therefore, no Condorcet-paradox issues arise from this process. An example of this voting procedure can be seen in Figure 3. The true value of the enacted policy platform can then be determined in the ground truth solution space.

Note that agents don't necessarily have to agree with each other's reasoning or even have the right reasoning in order to come to good decisions. Rather, each agent maintains their own reasons for favoring or disfavoring individual policies, and agents may ultimately find that they agree on some policy implementations while disagreeing on *why* they hold those preferences. This process of agents sharing and considering reasons continues until the optimal policy platform is enacted or after 5000 timesteps.

This procedure gives three metrics by which we can assess deliberative outcomes: the percentage of good policies enacted, the distance of the implemented policy platform from the optimal platform, and the percentage of agents in the largest coalition. For the first measure, a policy is considered "good" if its state matches the optimal state. Thus, this value will be 1 if the enacted policies match the optimal policy platform. In the second measure, we use the ground truth influence values to determine the true value of the enacted policy platform and measure the absolute negative distance of this value from the value that would have been obtained by enacting the optimal policy platform. This number is 0 if the group identified the optimal solution and negative otherwise. Finally, we look at the percentage of agents who share the largest policy platform coalition. That is, agents are considered to share a coalition if they support the exact same policy platform – i.e., vote for and against the same policies. A value of 1 here indicates that agents have reached full consensus. Note, this doesn't necessarily mean that agents have identified the true optimal policy platform, merely that they have come to agreement as to which platform to enact.

These metrics can be evaluated after any timestep in a single simulation. In order to highlight trends in deliberative outcomes, however, we focus here on the final state of simulations from a given noise level. For each noise level, we run 100 simulations that each last until the optimal policy platform is enacted or after 5000 timesteps. We then record the percent of good policies enacted at the end of the simulation, the distance of the final policy from optimal, and the size of the largest coalition. We consider the average of these values along with a bootstrapped 95% confidence interval.

Figure 3. Example policy vote.

Results

Effects of Cognitive Capacity and Ideological Skew

We first examine how different types of reasoners perform along various deliberative metrics. This includes uninformed agents, whose beliefs are evenly distributed around the ground truth, ideologues, whose beliefs are heavily skewed either positively or negatively from the ground truth, and intellectuals, whose beliefs are closely tied to the ground truth independent of noise value.

As we can see in Figure 4, groups composed entirely of uninformed agents perform increasingly poorly as the noise level increases. This is to be expected, as the noise level indicates that these agents are initialized with beliefs further and further from the ground truth. Since the ground truth influence values are initialized to be between −10 and 10, at even modest levels of noise, the beliefs of these uninformed agents are essentially random. Since intellectual agents are initiated independently of the noise level, we similarly see that such agents perform consistently well regardless of noise level. While in principle, these findings reinforce concerns about the influence of cognitive capacity on deliberative outcomes (Lippmann, 1922; Tversky & Kahneman, 1981), both uninformed and intellectual agents were designed to serve primarily as bounding conditions representing extremes of poor and good cognitive capacity. In other words, while these findings do confirm that ideal, "intellectual" agents would come to good solutions when deliberating, the failure of uninformed agents should not necessarily be interpreted as a failure of real-world citizens.

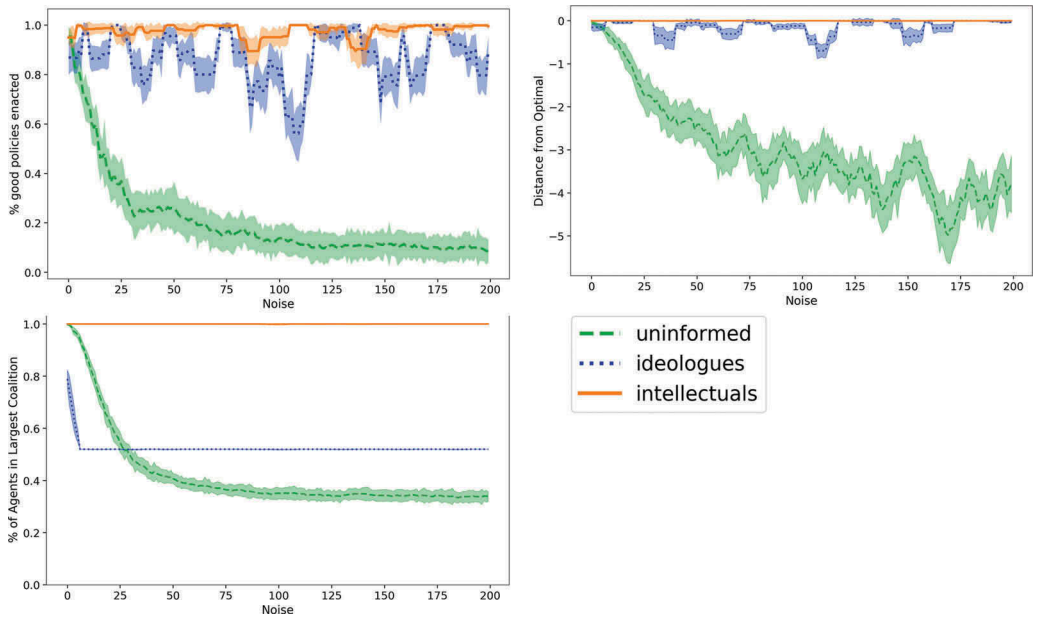

Figure 4. Deliberative outcomes by reasoning type.

Interestingly, factioned groups composed entirely of ideologues perform reasonably well over all noise levels, suggesting that the process of "giving and asking for reasons" can, in fact, help groups identify good solutions even when their initial information is poor. Because ideologues are intentionally skewed away from the ground truth, but in differing directions, their competing presence appears to bring some balance to the overall group, allowing some amount of the ground truth signal to remain. This finding suggests that even if humans are inherently poor at reasoning, it may actually be beneficial for our collective biases to be nonrandomly distributed. While this finding does not alleviate all fears that factions will be unable to successfully collaborate (Dworkin, 2006), it does ameliorate this concern for settings in which people are genuinely willing to consider other views. In other words, this finding supports core deliberative arguments, illustrating that even groups whose members are biased and cognitively flawed may be able to come to good solutions by genuinely trying to learn from and convince each other.

We more closely examine the influence of intellectuals on deliberative outcomes in Figure 5. Here, we consider both groups which have a majority of intellectuals as well as groups which have a majority of ideologues. In majority intellectual groups, half of the agents are intellectuals while the other half is split evenly between ideologues with positive and negative skew. In majority ideologue groups, only 1–2 agents are intellectuals, and the bulk of agents are split in ideological skew.

However, we find little evidence to suggest that intellectuals serve to improve deliberative outcomes. While a group composed entirely of intellectuals quickly identifies optimal solutions – as shown in Figure 4 – their presence does not seem to enhance the findings of ideologue agents. One reason for this may be found in the group's coalition behavior. As we would expect, groups composed entirely of ideologues form a 50% voting bloc, representing the polarization between the positively and negatively skewed ideologue groups. When a large number of intellectuals are added to the mix, the size of the largest coalition increases significantly – here, holding steady at around 80%. This suggests that while these groups converge to near-consensus, they do not often converge on the optimal solution. This could suggest that even a large group of intellectuals is not enough to successfully moderate polarized spaces, or it could be an effect of the model – similar

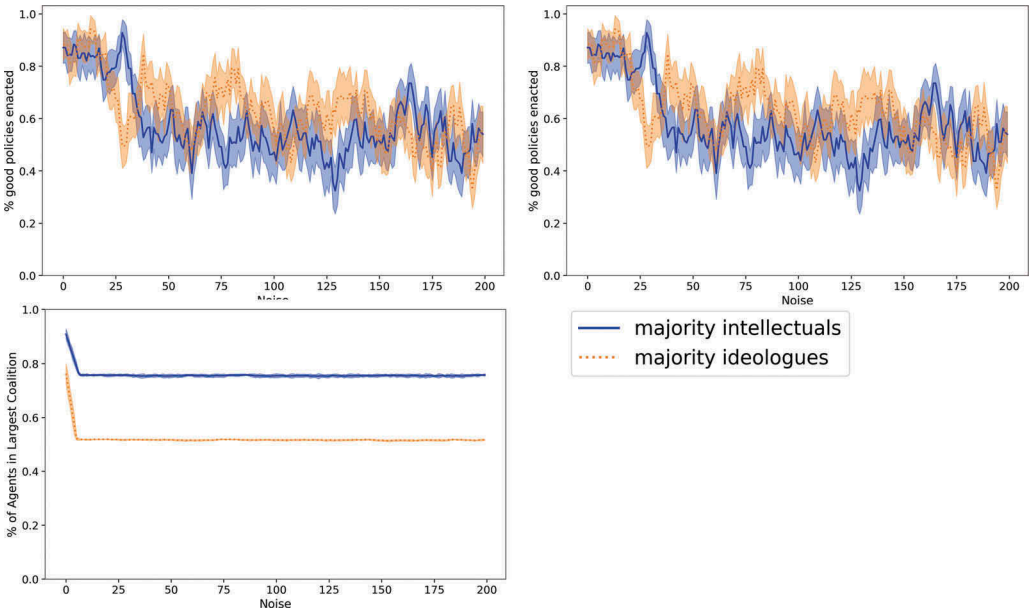

Figure 5. Deliberative outcomes portion of intellectuals.

to the finding of Lazer and Friedman (2007) that efficient systems tend to converge too quickly on sub-optimal solutions. While similar results have been found in other agent-based models (March, 1991), this result has not been replicated in human experiments (Mason & Watts, 2012), where moderating behavior may emerge differently than in simulated systems.

Effect of Cognitive Capacity and Open-mindedness

Finally, we examine the effects of agents' willingness to accept others' views, testing whether issues of group polarization (Sunstein, 2002) or confirmation bias (Nickerson, 1998) lead groups to reach worse deliberative outcomes. In Figure 6, we compare outcomes among groups of ideologues who are skeptical (only accept if the vectors have a cosine similarity greater than +.75), who are moderate (cosine must be positive) and who are open-minded (will accept if the vectors have a cosine similarity greater than −.75). Here, we see virtually no differences between the three groups. Indeed, skeptical and open agents appear to perform equally and, if anything, slightly outperform their moderate peers. More simulations would be needed to determine whether or not this

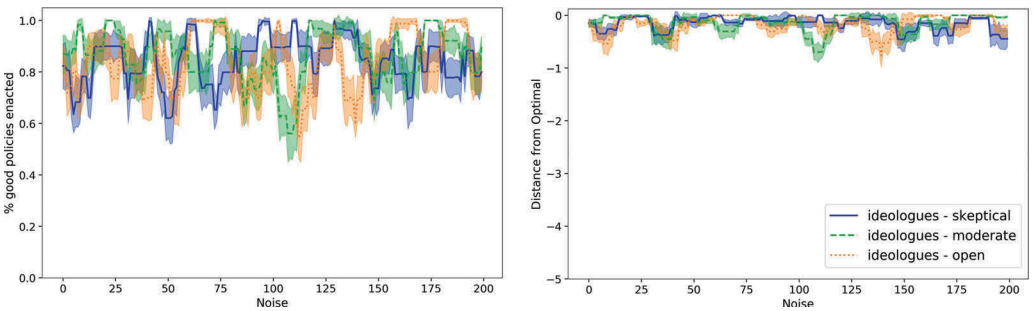

Figure 6. Deliberative outcomes by open-mindedness.

performance is a significant effect, but it may reflect a tendency for moderates to accept just the wrong amount of information – i.e., they are convinced to change their views, but frequently change their views to wrong things.

Alternatively, this finding may reflect the model's limiting assumption that all agents share the same "open-mindedness" parameter. A more complex model could further examine heterogeneities in accepting beliefs and would be more likely to uncover variations based on this property. For example, we might expect groups to come to better decisions when the agents with the most accurate beliefs are the least willing to abandon their views. By holding the tendency to accept others' views constant across the group, we see little gain or loss in the group's overall performance.

Discussion

Agent-based models provide a promising approach to exploring the dynamics of group problem-solving in normative settings. While necessarily highly stylized, such models allow for tunable control over parameters that could never be manipulated in real life. Furthermore, in breaking complex processes down into component parts, these models help us think differently about macroscopic phenomena – probing the constituent elements and gaining insight into the role that even simple mechanisms can play. Previous agent-based models of belief systems, for example, have provided insight into processes of group consensus and dissensus (Altafini, 2013; DeGroot, 1974; Friedkin et al., 2016).

The methodological approach presented here introduces a framework for using agent-based models to interrogate specific aspects of communication dynamics. By developing tunable parameters related to individual agents' predilections and biases, the model showcases how differing individual attributes can lead to notably different outcomes. While the substantive focus of this paper is on peer deliberation, the methodological framework is suitable for any process of communicative exchange. This includes formal debates, informal conversation, and the dissemination of information or misinformation. Coupling the methodological approach introduced here with models of spreading phenomena could be particularly powerful as a means to examine how individual tendencies to accept or share information could influence the spread of specific messages.

Substantively, this has focused on a key challenge of political communication – that, even when given access to the same information, people frequently do not agree as to the best course of action. Such normative settings are common in the real world and can be very high-stakes. This paper focuses on one particularly salient setting: examining group decisionmaking processes around policy implementation. The model imagines a small group of deliberators considering a set of possible policies all aimed at addressing some overarching social issue, such as education, healthcare, or crime. Agents will ultimately vote on which policies to implement but are given the opportunity to share information, reasons, and justification before they do.

While there are good reasons to be skeptical that such a deliberative process could meaningfully influence a vote in the real world, empirical evidence suggests that such influence is possible (Fishkin, 2014; Knobloch et al., 2013; Neblo et al., 2010). Furthermore, even if the deliberative ideal of reasoned exchange (Mansbridge, 2015) is not frequently met, the real-world implications and repercussions of such decision-making processes obligates us to try to understand the dynamics of these processes and to assess the conditions under which they lead to better outcomes.

Specifically, this paper examines three canonical deliberative failures: limited cognitive capacity, group factions, and poor judgment for accepting or rejecting views. Each dimension captures some element of real-world deliberative challenges. People may be less likely to come to good decisions if humans have limited cognitive capacity, if they are too polarized to engage with people unlike themselves, or if they are too set – or too flexible – in their opinions. I examine these effects through groups composed of uninformed, intellectual, or ideologue agents who operate with differing rules for adopting others' beliefs. I imagine small groups of deliberators attempting to engage in good-faith reasoning about topics on which they hold fundamentally differing views. Using the *NK* model to operationalize a complex policy

landscape, agents engage in a game of giving and asking for reasons – exchanging policy preferences as well as justification for those preferences.

The key finding is that polarized groups do surprisingly well at identifying optimal policy solutions. While uninformed agents perform as expected – achieving worse outcomes as cognitive capacity decreases – groups composed of oppositely-skewed ideologues appear to be resilient to the effects of declining cognition. This finding is in line with deliberative theory (Mansbridge, 1999) and suggests that heterogeneous agents can achieve good outcomes *if* they are able to engage in good-faith discussions.

This, of course, is a very restrictive assumption – very often, people do not engage in goodfaith exchange and are more concerned with winning than in collaboratively discovering the truth. In such settings, group factions would likely lead to entrenchment rather than optimal outcomes. What this finding suggests, though, is that the presence of opposing factions may not itself be the biggest concern. Indeed, having access to a diversity of opinions has the potential to lead to better outcomes than might be achieved otherwise. This benefit can only be realized, however, if we are able to build systems, institutions, and interventions which help people truly listen to and learn from each other.

We further find that this ideological effect does not appear to be moderated by the other parameters of interest. That is, while ideologues do far better than uninformed agents, their performance is not significantly improved by the presence of highly cognitive intellectuals or by being increasingly open to, or skeptical of, others' views. This suggests that the main factor in driving ideologues' performance is their inherent counter-balance to each other; by constantly pushing the "other side" to be better, both groups, and ultimately the whole, can improve.

It is particularly interesting to note that the conditions under which agents accept each other's views have little effect on observed outcomes. While we would see entrenchment if agents were wholly unwilling to consider opposing views, even a modest openness to others' opinions can lead to better outcomes.

Taken together, these results suggest a need to further invest in studying and creating deliberative spaces. The deliberative ideal is indeed lofty – and we should not expect that every person in every conversation will fulfill the ideals of good-faith reasoned exchange (Mansbridge, 1999). But these results suggest that we don't have to. A world in which discussion and debate lead to better policy outcomes doesn't have to be perfect – it doesn't have to rely on highly cognitive people wholly free of partisan bias. It is okay for people to be imperfect and to be flawed in their thinking – but *only* if they are modestly willing to listen to each other.

Disclosure Statement

No potential conflict of interest was reported by the author.

References

Altafini, C. (2013). Consensus problems on networks with antagonistic interactions. *IEEE Transactions on Automatic Control*, *58*(4), 935–946. https://doi.org/10.1109/TAC.2012.2224251

Choi, T., & Robertson, P. J. (2013). Deliberation and decision in collaborative governance: A simulation of approaches to mitigate power imbalance. *Journal of Public Administration Research and Theory*, *58*(4), 495–518. https://doi.org/10.1093/jopart/mut003

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, *69*(345), 118–121. https://doi.org/10.1080/01621459.1974.10480137

Dworkin, R. (2006). *Is democracy possible here?: Principles for a new political debate*. Princeton University Press.

Eliasoph, N. (1998). *Avoiding politics: How Americans produce apathy in everyday life*. Cambridge University Press.

Evans, J. S. B. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. https://doi.org/10.1016/j.tics.2003.08.012

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Fishkin, J. (2014). Reviving deliberative democracy. In J. Attali et al., *La Démocratie enrayée*?, Brussels: Académie royale, 181–200.

Friedkin, N. E., Proskurnikov, A. V., Tempo, R., & Parsegov, S. E. (2016). Network science on belief system dynamics under logic constraints. *Science*, *354*(6310), 321–326. https://doi.org/10.1126/science.aag2624

Gaventa, J. (1982). *Power and powerlessness: Quiescence and rebellion in an Appalachian valley*. University of Illinois Press.

Geisendorf, S. (2010). Searching nk fitness landscapes: On the trade off between speed and quality in complex problem solving. *Computational Economics*, *35*(4), 395–406. https://doi.org/10.1007/s10614-009-9192-4

Gutmann, A., & Thompson, D. F. (1998). *Democracy and disagreement*. Harvard University Press.

Herrmann, S., Grahl, J., & Rothlauf, F. (2014). Problem complexity in parallel problem solving. In *Proceedings of the International Workshops SOCNET and FGNET*, Bamberg, Germany, 77–83.

Hibbing, J., & Theiss-Morse, E. (2002). *Stealth democracy: Americans beliefs about how Government should work*. Cambridge University Press.

Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes*. Houghton Mifflin.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kauffman, S., & Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, *128*(1), 11–45. https://doi.org/10.1016/S0022-5193(87)80029-2

Kauffman, S. A., & Weinberger, E. D. (1989). The nk model of rugged fitness landscapes and its application to maturation of the immune response. *Journal of Theoretical Biology*, *141*(2), 211–245. https://doi.org/10.1016/S0022-5193(89)80019-0

Knobloch, K. R., Gastil, J., Reedy, J., & Cramer Walsh, K. (2013). Did they deliberate? Applying an evaluative model of democratic deliberation to the oregon citizens' initiative review. *Journal of Applied Communication Research*, *41*(2), 105–125. https://doi.org/10.1080/00909882.2012.760746

Lazer, D., & Friedman, A. (2007). The network structure of exploration and exploitation. *Administrative Science Quarterly*, *52*(4), 667–694. https://doi.org/10.2189/asqu.52.4.667

Levinthal, D. A. (1997). Adaptation on rugged landscapes. *Management Science*, *43*(7), 934–950. https://doi.org/10.1287/mnsc.43.7.934

Lippmann, W. (1922). *Public opinion*. Harcourt, Brace and Co.

Madison, J. (1787). Federalist no. 10: "the same subject continued: The union as a safeguard against domestic faction and insurrection". *New York Daily Advertiser*.

Manin, B. (2005). *Democratic deliberation: Why we should promote debate rather than discussion*. Paper delivered at the Program in Ethics and Public Affairs Seminar, Princeton University. https://as.nyu.edu/content/dam/nyu-as/faculty/documents/delib.pdf

Mansbridge, J. (1999). Everyday talk in the deliberative system. In S. Macedo (Ed.), *Deliberative politics: Essays on democracy and disagreement* (pp. 1–211). Oxford University Press.

Mansbridge, J. (2015). A minimalist definition of deliberation. In Patrick Heller and Vijayendra Rao (Eds.), *Deliberation and development: Rethinking the role of voice and collective action in unequal societies* (pp. 27–50). World Bank.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–87. https://doi.org/10.1287/orsc.2.1.71

Mason, W., & Watts, D. J. (2012). Collaborative learning in networks. *Proceedings of the National Academy of Sciences*, *109*(3), 764–769. https://doi.org/10.1073/pnas.1110069108

Mercier, H., & Landemore, H. (2012). Reasoning is for arguing: Understanding the successes and failures of deliberation. *Political Psychology*, *33*(2), 243–258. https://doi.org/10.1111/j.1467-9221.2012.00873.x

Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press.

Neblo, M. A. (2015). *Deliberative democracy between theory and practice*. Oxford University Press.

Neblo, M. A., Esterling, K. M., Kennedy, R. P., Lazer, D. M., & Sokhey, A. E. (2010). Who wants to deliberate—and why? *American Political Science Review*, *104*(3), 566–583. https://doi.org/10.1017/S0003055410000298

Neblo, M. A., Esterling, K. M., & Lazer, D. M. (2018). *Politics with the people: Building a directly representative democracy* (Vol. 555). Cambridge University Press.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. In R. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* (pp. 79–96). Psychology Press.

Sanders, L. M. (1997). Against deliberation. *Political Theory*, *25*(3), 347–376. https://doi.org/10.1177/0090591797025003002

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, *134*(2), 207. https://doi.org/10.1037/0033-2909.134.2.207

Shore, J., Bernstein, E., & Lazer, D. (2015). Facts and figuring: An experimental investigation of network structure and performance in information and solution spaces. *Organization Science*, *26*(5), 1432–1446. https://doi.org/10.1287/orsc.2015.0980

Sunstein, C. R. (2002). The law of group polarization. *Journal of Political Philosophy*, *10*(2), 175–195. https://doi.org/10.1111/1467-9760.00148
Sunstein, C. R. (2009). *Going to extremes: How like minds unite and divide*. University Press.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683