The Structure of Reasoning: Inferring Conceptual Networks from Short Text

Sarah Shugars[1]

[1]Network Science Institute, Northeastern University

April 22, 2019

Abstract

Behavioral models of political engagement typically neglect the structure of human reasoning, assuming instead that opinions represent random samples from some collection of retained information. Yet, scholarship in a number of fields has long indicated that cognitive processes as diverse as reasoning, arguing, remembering, and learning are best modeled as conceptual networks in which connections between similar ideas facilitate the storage and retrieval of relevant information. This structural dimension of reasoning has the potential to significantly influence how an individual samples from and acts on their available beliefs – some people may be prone to constantly return to one central idea, others may jump freely from topic to topic, and others may struggle to see how an issue is relevant to their interests all. This suggests that models of political behavior need to better integrate cognitive models of individual-level reasoning structure. What types of personality traits lead to what types of reasoning structures? How might a tendency towards different structures influence political behavior? Informed by work in political behavior and psychology, this project presents a generative model of individual reasoning in which latent personality traits encourage the activation of different reasoning structures. Using three datasets, spanning a detailed personality study, a nationally-representative poll, and political conversations on Twitter, I demonstrate that individual reasoning structure can be meaningfully inferred from short text and find that these structures correlate with validated personality and ideology measures. Ultimately, this work presents a collection of archetypes of individual reasoning which represent different moral philosophy stances and serve to better inform our understanding of political behavior.

1 Introduction

Public opinion is the bedrock of the the democratic ideal, yet it is often dismissed as the incoherent whims of the masses. "The people" are far too prone to bias, too easily misled, and too set in their ways to effectively reason together. While there is strong evidence documenting each of these concerns, broad distrust of "the people" creates a fundamental democratic challenge: how can a society responsibly enact the will of the people while protecting against factions, populism, and the tyranny of the majority? Designing democratic institutions capable of addressing this challenge therefore requires an understanding not only of *what* the people think, but rather *why* they think it; it requires behavioral models capable of examining individual processes of opinion formation and change.

Existing models of political behavior, however, are insufficient to capture the dynamics and complexities of our modern world. Social influence models in which media and key gatekeepers (Lazarsfeld, 1948; V. O. Key, 1942; Zaller et al., 1992) craft the available narratives of public opinion are difficult to operationalize in a world of viral tweets and fleeting internet stars. Party identification models (Campbell, 1960) have been bolstered by increasing levels of affective polarization but are unable to address the realities of fracturing parties and tea-party politics (Blum, 2019). Finally, psychological models (Haidt and Joseph, 2008; John and Srivastava, 1999), which bring the promise of incorporating cognitive complexity, fail to consider the broader social and informational context in which an individual's cognitive processes are taking place. Taken together, the existing literature provides a rich but incomplete starting point for understanding modern processes of public opinion.

Furthermore, in focusing on the *content* of opinions, these models neglect the growing literature documenting the significance of the *structure* of reasoning. That is, they take an "opinion" to be an atomistic entity which can be received, retrieved, and transmitted with varying fidelity and focus their attention on the conditions which improve or degrade those processes. Work in a number of fields, however, has found that reasoning is more than a collection of unstructured ideas; people store and retrieve information not as isolated packets of information, but as complex networks of interconnected concepts. Studies of reasoning (Axelrod, 1976; Carley, 1993; Toulmin, 1958), arguing (Toulmin, 1958; Walton, 1996), remembering (Collins and Loftus, 1975; Quillian, 1967), and learning (Shaffer et al., 2009; Shavelson, 1974) all suggest that individuals express and interpret beliefs through network structures. When speaking with others, we raise ideas that seem related to what they said; when thinking to ourselves, we move from idea to idea via their connections; and when assessing a complex issue, we weigh the pros and cons as well as their interconnections in order to arrive at a final judgment. Network interpretations of the cognitive organization of knowledge are bolstered by behavioral observation of arguments, deliberation, written texts, and self-reports that repeatedly suggest that individuals perceive their ideas to be connected to each other in complex networks of support or contradiction.

Incorporating a structural dimension to our understanding of political reasoning holds the potential to better explain the joint social and cognitive processes at play in the formation of public opinion; processes which are critical for identifying and ameliorating democratic woes. While we may scoff at the deliberative ideal of citizens with differing opinions, experience, and information successfully reasoning together about matters of mutual concern (Cohen, 1989; Habermas, 1984; Rawls, 1971; Mansbridge, 2015), political talk – even if uninformed or antagonistic – is a notable part of our public sphere. Modeling deliberative exchange, casual kitchen-table conversations, or even online fights, however, implicitly suggests conceptual structures that go far beyond atomic preferences. Indeed such models rely upon a network structure where ideas are connected to each other by implication or association, and where interlocutors select arguments based on the inferred effect those points will have on the interconnected beliefs of the other participants.

In other words, political disagreements center not only on what people say, but fundamentally on how they say it. Misinformation thrives on framing which magnifies prior beliefs. Polarization is perpetuated by disagreement about basic facts and their relations. Deliberative theory requires far more than raw communication and information exchange between citizens; it requires citizens who are capable of connecting their experiences to one another (Habermas, 1984; Dewey, 1927).

Yet this structural dimension of political reasoning is not well understood and a network model cannot emerge from the current behavioral methods which ask about disconnected issues (Haste, 2013). How do behavioral traits influence the ways in which individuals structure their thinking? Do liberals and conservatives think in fundamentally different ways? Do demographic factors influence not just what we think, but *how* we think? Can people who structure their thinking in different ways find common ground? Can they even begin to communicate with each other?

In this paper, I present a typology of expected reasoning structures as well as a new method for inferring these structures from short text. I translate particularist and utilitarian moral philosophies into structural dimensions of connectivity, complexity, and hierarchy, and present methods for measuring these as network features. Leveraging the grammatical structure of language, I present a method for inferring the structure of individual-level reasoning from short text. I apply this method to three distinct datasets, spanning a detailed personality study, a nationally-representative poll, and political conversations on Twitter. I demonstrate that individual reasoning structure can be meaningfully inferred from short text and find that these structures correlate with validated personality and ideology measures.

2 Related Work

Public opinion has long been modeled as a top-down process in which individual political opinions are largely inherited from media, opinion leaders, political parties, or interest groups (Lippmann, 1922; Lazarsfeld, 1948; Campbell, 1960; V. O. Key, 1942). More recently, Zaller et al. (1992) introduced a psychological dimension to these models, arguing that individuals do not hold static preferences, but rather sample from available information when prompted by a survey battery. In this model, elite discourse still shapes the dominate narratives an individual is exposed to, but the information that individual retains is mediated by behavioral factors and the position they espouse on a survey is little more than a random draw from their pool of available, relevant information. In this sense, "public opinion" as measured by survey research is better interpreted as a noisy signal of elite discourse rather than as a reflection of an individual's true preferences or beliefs.

My work here draws heavily on Zaller, assuming that individuals do not have well-developed, fixed positions on the majority of political issues. Rather, when asked to explain their position – either on a survey or in a conversation – they sample from an individual conceptual network of ideas they consider to be linked to the topic at hand. While the stability of the specific concepts included in this retrieved network structure may vary with the salience of the specific issue, the structure of that network has the potential to reveal behavioral traits and suggest the degree to which an individual is prepared to engage in productive political discussion on the topic.

Psychologists have further argued that individual opinions are not driven by reason at all, but rather by clusters of latent traits (Haidt and Joseph, 2008; Haidt, 2012; John and Srivastava, 1999) which are usually opaque to the subjects (Graham et al., 2011). Under these models, individual judgement is made almost exclusively by gut feeling alone and any articulated reasoning is merely the result of post-rationalization. In the political realm, this often translates to individuals attaching themselves to the people, parties, or interest groups which best appeal to their latent personality traits and then simply toeing the party line on specific policy issues. This line of research has yielded persuasive results, yet risks taking too myopic a view of human behavior. First, while people may indeed be more prone to the easy path of "thinking fast" (Kahneman, 2011) and engaging in the rapid and instinctual cognition of "System 1" thinking (Evans, 2003), the story of human cognition is not complete without an understanding of the complementary processes of "thinking slow" (Kahneman, 2011) and engaging rational "System 2" thinking (Evans, 2003). Second, people *do* regularly express their beliefs as interconnected networks of ideas (Axelrod, 1976; Shaffer et al., 2009; Shavelson, 1974), making these structures an important piece of political engagement whether you believe them to be the result of explicit reasoning or mere grammatical formality.

Conceptual networks have been explored across a variety of disciplines under varying names including semantic networks, knowledge structures, mental maps, and epistemic networks. The core intuition underlying most approaches is that human thought is best represented as networks of interconnected concepts (Collins and Loftus, 1975; Dorsey et al., 1999; Quillian, 1967), but because these concepts cannot be directly observed, they must be inferred primarily through language. This inference process has generally proceeded from two directions: a psychological approach, which begins with theories of cognition and memory and attempts to recover these structures through experimentation, observed behavior, and collaborative knowledge-building; and a linguistic or logical approach, which seeks to explain linguistic patterns, meanings, and grammars using inferred network structures. These two strains of study often converge on a similar set of conceptual models, but reflect the varied disciplines targeting a shared problem. An additional stream of work in moral philosophy has aimed to normatively assess the quality of an individual's conceptual network, bracketing the question of measurement itself.

On the psychological side, Quillian's theory of semantic memory argues that human memory search is made possible by storing information as a network in which concepts, represented as nodes, are connected by relational links to other conceptual nodes (Collins and Loftus, 1975; Quillian, 1967). In this model, each node provides a shallow understanding of a given concept and is represented by a single word or phrase. A "concept" more deeply considered, however, contains indefinitely large amounts of information and is more properly expressed as the entire network accessible from a given concept node (Collins and Loftus, 1975). Such a knowledge structure allows a person store a concept as a compressed object (node) while simultaneously allowing access to a richer understanding through the network structure (Quillian, 1967). This is the core intuition behind semantic network libraries such as BabelNet (Navigli and Ponzetto, 2012), ConceptNet (Speer and Havasi, 2012), and SNePS (Shapiro and Rapaport, 1987): a concept, encoded as a word, can be best described though its associated concepts, which themselves are encoded as words. The end result is a semantic network of connected words representing an underlying conceptual network of connected concepts.

Scholars in numerous fields have developed models of conceptual networks in applied contexts. Within the educational literature, cognitive models suggest a framework for evaluating learning. If knowledge itself has a network structure, then mastery cannot properly be measured by the ability to repeat a list of facts. Rather, students must learn the relevant knowledge structures, building an understanding not just of key concepts, but of the connections between those concepts. By understanding how relevant information is interconnected, students develop the ability to apply existing knowledge to new situations (Dorsey et al., 1999; Hong et al., 2004; Shaffer et al., 2009; Shavelson, 1974). Social scientists have similarly argued that conceptual networks can be used to examine how individuals reason and make choices between alternatives (Axelrod, 1976; Carley, 1993). In weighing possible outcomes, a person evaluates connected concepts and consequences; exploring paths within their conceptual network in order to determine the optimal choice. Deliberation provides a natural venue to extend such models, as participants may enter conversation with differing views and must therefore attempt to share structured knowledge before reaching a decision. Notably, the communication of such structured knowledge must be done through language and is influenced by the structural features of language (Axelrod, 1976; Eveland and Cortese, 2004).

The second major approach is via linguistic and/or logical models, which study the structure of language as a proxy for the structure of knowledge. Perhaps the most well developed such models trace their roots back to Aristotelian efforts to define the structure of argumentation (Toulmin, 1958). Such structures may be relatively simple: a major premise connected to a minor premise leads inevitably to a logical conclusion; or it may be significantly more complex, such as in the two dozen schemes described by Walton (1996) or the Context Free Grammar introduced by Mochales and Moens (2011). But while theorists have differed in the specifics of the models they put forth, their approaches all begin with implicit acceptance of the network structure of arguments: the soundness of a conclusion rests not only upon the ideas supporting it, but on the ways in which those ideas are connected. In other words, arguments fundamentally have a coherent structure expressed through linguistic structure and defined by evidence relationships (Cohen, 1987). The search for these structures has given rise to a rich body of research known as argument mining, in which supervised and semi-supervised computational methods automate the search for the sorts of argument structures articulated by Aristotle or Toulmin (Mochales and Moens, 2011). The conceptual networks inferred via these methods tend to be more structured and hierarchical than those inferred from open-ended psychological approaches, but the basic structure of nodes and edges representing ideas and their interconnections remains.

While psychological and linguistic approaches aim to examine the structure of conceptual networks, an important stream of work in philosophy has developed normative theories around the properties of these networks. Many philosophers are committed to coherence, considering a moral position valid insofar as it is coherent with other views. What constitutes "coherence," however, differs between philosophers, leading to differing topological interpretations. For instance, in Henry Sidgwick's influential version of utilitarianism, "the current moral rules" such as "do not lie" are used to generate most of our actual judgments (Sidgwick, 1907). The principle of utility, however, serves as a gatekeeper through its connection with these moral rules, so it has an importance in his system commensurate with the network concept of high betweenness centrality. In particularist moral theories, by contrast, each moral judgment is only linked to others by loose and local analogies (Dancy, 1993), implying that no ideas should enjoy disproportionate centrality in a person's whole network of moral ideas. McNaughton and Rawling (2000) argue that this is the flaw of particularism, for some concepts really are "central" to morality. Another approach is to consider the coherence of a set of beliefs, which Rawls (1993) asserts is a condition of rationality. He defines "a reasonable doctrine" as "an exercise of theoretical reason...It organizes and characterizes recognized values so that they are compatible with each other and express an intelligible view of the world." More broadly, several authors—e.g., Christen and Ott (2013); Dorsey (2006)—argue that coherence is an important indicator of validity. While there is no shared definition of what "coherence" looks like, the core argument in these theories is that consistency between individual pairs of beliefs is too low a standard – since beliefs can be consistent but completely unrelated – but expecting all pairs of beliefs to be directly connected is too stringent a standard – because moral views range over a wide variety of topics. Thagard (1998) proposes a theory of coherence that involves literal network relations, but he overlooks many of the relevant formal features of networks, in part because his examples are very small sets of related ideas. Berker (2015) posits that an individual's beliefs should be modeled as a network to reveal its degree of coherence and begins to explore the variety of forms that a network of moral values can take.

In this paper, I operationalize these differing normative views through three dimensions of network characteristics: connectivity, complexity, and hierarchy. Connectivity here serves as a baseline for the principal of coherence which is shared across normative approaches. Complexity reflects the particularist view (Dancy, 1993) which emphasizes the importance of adaptive, context-dependent reasoning. Finally, hierarchy represents the utilitarian view (Sidgwick, 1907), in which select ideas serve as core guiding principles. I describe the specific measurement for each of these dimensions in detail in Section 3. While this work draws upon the normative judgements of moral philosophy, I do not aim here to proscribe normative weight to these various network structures. Rather, I seek to present a method for inferring the existence of these structures and evaluating the degree to which these structures correlate with known psychological and behavioral traits. In future work, I plan to assess these structural arrangements in terms of how conducive they are to productive political talk.

3 Model and Network Measures

If individual conceptual network structure is to serve as a useful tool for deepening our understanding of public opinion, then we must start with a strategy for interpreting the structure of an inferred network. While I describe my method for inferring these structures from text in Section 4, I begin here with a generative network model of the reasoning processes, followed by a description of the dimensions along which I assess inferred network structure and the network measures which indicate those dimensions.

3.1 Received-Accepted-Network Sample

I build here off Zaller's Received-Accepted-Sample model (Zaller et al., 1992), focusing particularly on the *sample* stage in which an individual expresses a political belief. In Zaller et al. (1992)'s original model, individuals receive information from elites, selectively accept information which conforms to prior beliefs, and then, when prompted by a pollster, generate a Likert-scale position on the fly by sampling from their accepted beliefs. While interrogating the process of accepting information is beyond the scope of the current study, I here follow the semantic memory literature (Collins and Loftus, 1975; Quillian, 1967) and assume that any accepted information is stored cognitively as a network structure. I then take "sampling" to be a network sampling process; e.g., when prompted to express their reasoning, either by a pollster or in conversation, individuals sample salient nodes from this latent conceptual network and describe the ideas they see as connected to those nodes. This is not to suggest that individuals are walking around with fully-formed and established networks in their heads. Rather the "accepted" network exists as an amorphous latent structure that can be better interpreted as a quiescent collection of heuristics rather than as a fixed view itself. I would posit that the salience of a given issue and the frequency with which someone expresses their opinions on that issue would effect the stability of a person's "accepted" network structure, but such explorations are left for future work.

More formally, given some accepted network A comprised of nodes N and edges E, a person will express their opinion as a sampled network $S \subset A$. They will do this by first sampling N for the most salient concepts, retrieving one or more starting nodes $n = \{n_1, \dots, n_i\} \in N$. They will then search E for relationships involving these key concepts, retrieving some collection $e = \{(n_i, n_j), \dots\} \in E$ where at least one element of each retrieved edge is in n. These relations then bring new concepts into the conversation, which in turn may lead to the retrieval of additional relationships. This process is repeated iteratively until the subject gets tired, feels they have expressed all relevant, related ideas, or simply decides they have said enough to get their point across or otherwise satisfy their interlocutor. Given limits on cognitive capacity and the nebulous nature of A, we would expect these processes to be noisy, with subjects neglecting to include relevant edges or nodes with some probability p.

This model does not put limitations on the precise order in which sampling occurs, and rather takes this to be a behavioral trait with the potential to influence the ultimate structure of S. Some subjects may choose to express their ideas by beginning with one concept and following that concept's connections as far as they can (i.e, a depth-first-search), while others might start with a few concepts and reason evenly from each (i.e., a breadth-first-search). In this paper, I focus solely on individual variation in resulting network structure, leaving the influence of sampling strategies for future work.

Once constructed through this sampling process, S is then expressed as observable semantic output with a latent conceptual structure governed by the network structure of S. While I argue here that the structure of S is tied to behavioral traits and represents different approaches to reasoning, it is important to note that the exact mechanism of this reasoning falls beyond the scope of this current work. That is, the structure of S could plausibly represent either an individual's true internal reasoning process or simply their external sense of what makes a good argument – and this mechanism need not be consistent across individuals. Regardless of which explanation you favor, however, expressing a view itself is a meaningful political act (Austin, 1962) with important implications for deliberation, discourse, and civic health. While it perhaps does not matter how a subject expresses themselves to a pollster, whether they are sharing their own reasoning or repeating arguments they have heard from others, it *does* matter how they express their political views to friends, family, and strangers on the internet.

3.2 Connectivity, Complexity, Hierarchy

In this paper, I examine the extent to which the structure of individuals' expressed reasoning is consistent with normative approaches of moral philosophy. Specifically, after using course semantic features to infer the structure of a subject's sampled network S, as described in Section 4, I examine the inferred network along three dimensions: connectivity, complexity, and hierarchy. Connectivity provides a baseline measure of coherence, while complexity reflects the particularist philosophy and hierarchy represents the utilitarian view. Each of these dimensions and their related network measures are summarized in Table 1 and described in detail below.

Connectivity

One of the few things on which moral philosophers agree is that "coherence" is critical to the validity of a moral perspective (Dorsey, 2006; Christen and Ott, 2013). While we will see differing topological interpretations of coherence in the dimensions of complexity and hierarchy, I take connectivity to be the baseline for coherent thought. That is, for a position to be coherent, it must, at a minimum, connect all relevant points. These connections need not be positive in value and do not need to exist between every possible pair of ideas, but, if my position on an issue is coherent I ought to be able to describe any portion of my position in relation to any other portion of my position.

Topologically, I measure this connectivity as N_G/N, the fraction of nodes in the giant component, e.g., the largest set of connected nodes. If it is possible to get from any idea (node) to any other idea, then this ratio will equal 1, if no ideas are connected, it will equal $1/N$. This measure therefore captures the degree to which an individual is united or divided in their thinking – a subject whose reasoning is coherent would have a connectivity score of 1, while a subject who espouses disparate, unconnected views would have a positive score strictly < 1. Again, this connectivity measure should be taken as a baseline and does not fully represent all conceptions of coherence.

Complexity

Particularist theories argue that moral judgements are inherently context-dependent and therefore our ability to form appropriate judgements must be context-dependent as well (Dancy, 1993). This suggests that no single principle ought to enjoy an outweighed place in the ideal reasoning process – rather, many ideas should coexist in a flexible and loosely connected web. When confronted with a moral question grounded in a particular, detailed context, a reasoner can then navigate this versatile collection of ideas to determine the best judgement for the given situation.

I operationalize this concept through two topological measures: density and entropy. Density measures the ratio of observed edges to possible edges and, in an undirected network, is calculated as $2E/(N \times (N-1)$, where E is the number of edges and N is the number of nodes. This calculation assumes that every edge connects two nodes and that there are no self-loops, e.g., at a maximum as single node can connect to $N-1$ other nodes. A density of 1 then represents a fully connected network where every node is connected to every other node and a density of 0 represents a network with no edges at all. This therefore captures a network's flexibility in engaging or avoiding any particular nodes, and thus captures the particularist conception of coherence.

I further use entropy as a measure of a the inferred network's complexity. Entropy estimates the amount of information contained in p_k, the network's normalized degree distribution (Shannon, 1948). This distribution gives the probability that a randomly selected node will have degree k, e.g., connections to k other nodes. The entropy of that distribution then captures how homogeneous or heterogeneous a distribution is. Calculated as $-\sum(p_k \times \log(p_k))$, networks with more nodes will have higher entropy and networks with more heterogeneous degree sequences will also have high entropy. I normalize each network's entropy by the maximum possible entropy given the network's number of nodes N, giving an entropy score between 0 (low entropy) and 1 (maximum entropy).

In order to calculate the complexity of a given network, I average the density and normalized entropy to give a complexity score between 0 and 1.

Hierarchy

While particularist theories emphasize the importance of flexibility in moral reasoning, utilitarian views argue that all moral judgements ought to be made through a core set of guiding principles (Sidgwick, 1907). This approach argues that contextual details should never override central moral tenants - i.e., murder is not morally justifiable under any circumstances.

This suggests a hierarchical network structure, where some ideas are more central while others are more peripheral. I operationalize this dimension by measuring k_{std}, the standard deviation of a network's degree sequence as well as the network's assortativity, r. Note that while there are numerous measures of node centrality, this does not make for a good comparative topological statistic, that is - while it is meaningful to compare the centrality of nodes within the same network, it is not meaningful to compare centrality scores of nodes taken from distinct networks.

The standard deviation of a network's degree distribution, k_{std}, indicates the heterogeneity of a network's connectivity. If all nodes have the same number of connections, k_{std} will be 0. Because I am interested here in the hierarchical structure of the network, I normalize the observed standard deviation against the maximum possible for a hub-and-spoke network with the same number of nodes. That is, for an observed network with N nodes, I compare k_{std} to the standard deviation obtained from a network that has one node with $N-1$ connections (hub) and $N-1$ "spoke" nodes with 1 connection each to the hub.

Assortativity, r, is measured as the Pearson correlation coefficient of a network's degree sequence and captures the degree homophily of the network – e.g., the tendency of nodes of degree k to connect to nodes of similar or dissimilar degree (Newman, 2003). I renormalize assortativity to have the same dimension and valence as standard deviation; defining a hub-and-spoke network to have a normalized assortativity of 1. Finally, I calculate a network's hierarchy score as the average between k_{std} and the renormalized value of r.

Basic measures	
Nodes	The number of nodes in a network, N. This can be considered roughly as the number of concepts a person brings up in response to a single-issue prompt.
Edges	The number of edges, E in the network. This can be considered as a measure of how many interconnections a person sees between the concepts they surface. In this paper, all edges are considered to be weighted, signed, and undirected. That is, two concepts can be heavily or weakly connected, and positive or negatively connected. Those connections are taken to be undirectional in that node A has the same relationship to B as node B has to A.
Connectivity indicators	
Giant component percent	The percent of nodes in the largest component of the network, N_G/N. This measure indicates how cohesive the network is. A value of 1 indicates the network has a single component (e.g., a path exists between any two nodes), while lower values indicate that the network has multiple, disconnected components.
Complexity indicators	
Density	The ratio of existing edges to the total possible edges, $2E/(N(N-1))$. This is a measures the overall interconnectivity of a network with a value of 1 indicating that every idea is connected to every other idea and a value of 0 indicated that no concepts (nodes) are connected.

Entropy	Estimates the amount of information contained in the network's normalized degree distribution (p_k) (Shannon, 1948). This measure is dependent on both the length of the distribution (eg, N) and the heterogeneity of the distribution. Measured as $-\sum(p_k \times \log(p_k))$
Hierarchy indicators	
Standard deviation of degree	The standard deviation of the network's degree distribution. Lower numbers indicate that nodes are more homogeneous in their degree whereas larger values indicate greater difference between the lowest and highest degree nodes in the network. For the purposes of this study, k_{std} is normalized against a hub-and-spoke network with the same number of nodes N.
Assortativity	Measured as the Pearson correlation coefficient, r, assortativity captures the degree homophily of the network (Newman, 2003). For this study, assortativity is renormalized to have the same dimension and valence as standard deviation. This gives a hub-and-spoke network, which would typically have $r = -1$, a normalized assortativity of $+1$.

Table 1: Measures of network structure.

4 Methods

If subjects do indeed engage in political conversation by sampling from a latent network of accepted, interconnected beliefs as described in Section 3, then any observed semantic output, i.e., written or verbal response, can be used to infer the structure of that subject's sampled network S. The intuition here is that when seeking to explain or justify a view, a person first reaches for salient concepts and then describes the interconnections between those concepts using natural language. Through this process of articulating one's reasoning, the network structure of S then manifests as coarse semantic features. This intuition relies upon language itself being structured by nature (Toulmin, 1958; Walton, 1996), and existing for the purpose of communication. That is, if we assume that a person is genuinely aiming to communicate their reasoning to another, then we – just as any listener – should, in principle, be able to reconstruct the structure of their core arguments (Abbott et al., 2011; Habernal and Gurevych, 2015).

Previous efforts to infer individual level conceptual networks have leveraged this intuition, inferring network structure by hand coding essays or transcripts and using term co-occurrences (Axelrod, 1976; Shaffer et al., 2009; Shavelson, 1974). Due to the challenges and limitations of this manual approach, however, the majority of large-scale efforts have been at the collective rather than individual level; asking participants to collectively map their shared knowledge (Atteveldt, 2017; Speer and Havasi, 2012; Navigli and Ponzetto, 2012).

This paper draws upon both these past streams of work, aiming to apply computational approaches to infer the argumentative structure of short, individually-generated text. Existing work in the area of argument mining has found this reconstruction can be done with high accuracy for well-formed arguments (Katzav and Reed, 2008; Mochales and Moens, 2011; Feng and Hirst, 2011) and can perform reasonably well on poorly-structured arguments which are difficult for humans to interpret as well (Habernal and Gurevych, 2015; Stab and Gurevych, 2014).

Leveraging the inherent structure of language, I infer the underlying network structure S of a short text T through the following steps, which I summarize here and describe in detail below:

1. Determine the dependency parse tree for each sentence in T
2. Lemmatize all words to their base form and merge repeating words to a single node
3. Remove stopwords and select dependency relations, maintaining the network structure of the parse and recording any negative relationships
4. Merge all sentence-level parsings into a single network, allowing edges to be weighted by co-occurrence

1. Dependency parsing

Any sentence can be parsed into a tree of dependencies. This structure determines the grammatical "root" of a sentence – typically the primary verb – and defines each other word in terms of how it

Word	Head	Dependency relation
Abortions	be	nominal subject
should	be	auxiliary
be	ROOT	ROOT
legal	be	adjectival complement
under	legal	preposition
certain	circumstances	adjectival modifier
circumstances	under	object
.	be	punctuation

Table 2: Dependency parse for sentence “Abortions should be legal under certain circumstances.”.

modifies the root word or one of its children. In this parsing, every word has a head (parent) and a relationship describing how it modifies that head. These dependency relations are more fine-grained than part-of-speech tagging, describing, for example, whether a given word is the subject or object of the sentence.

Because dependency parsing grammatically operates on the sentence-level, a unique parse is determined for each sentence in an individual’s corpora. These sentence-level parsing are then merged into a single network structure in step (4).

2. Lemmatization

Lemmatization uses morphological analysis to returns the base version of a word. For example, “are,” “is,” and “am” – all version of the infinitive verb “be” will be returned simply as “be”. The process similarly reduces plurals and possessives to their base form. While stemming is also a common approach for this task, lemmatization is the more sophisticated approach, judging the base form from morphological content rather than simply, for example, removing the conjugated part of a verb. During this step, I also reduce all words to lower case.

After lemmatization, any words which are repeated are taken to represent the same concept and to therefore be the same node. At this point, the dependency parse may no longer have a tree-like structure as any word can have a number of parents equal to the number of times it appears in the text.

3. Remove stopwords

In the next step I remove stopwords and select dependency relations, namely punctuation, symbols, determiners (e.g. "the") and auxiliary words (e.g. "should"). While this step helps remove grammatical clutter from a sentence, it should not be used to modify the underlying network structure. Therefore, if a word which has children is removed in this step, those become connected to each other, and, if applicable, are adopted by their grandparents, i.e., connected to the parent of the removed word. Negation words (e.g. "not") are also removed in this step, and any resulting edges between a grandparent and grandchild are taken to have negative weight.

In the example sentence shown in Table 2, the words "should", "be", "under", and the punctuation "." are all removed. The words "abortion" and "legal" are both children of the removed word "be" and are taken to be connected. "Circumstances" is the only child of the removed word "under" and becomes connected to it's grandparent "legal." If the sentence read "Abortions should not be legal..." then the remaining nodes of "abortion" and "legal" would have a negative edge between them.

4. Merge parsings

This step merges the individual sentence-level network parsings into a single network, again, taking words which occur multiple times to be the same node. The weight between any two nodes is then taken to be the sum of their weights across sentences. That is, if the words "abortion" and "legal" are linked in two of the parsed sentences, the final network would give this relationship a weight of 2. This captures the semantic emphasis of repetition and indicates if a subject uses the same wording multiple times.

5 Data and Hypotheses

I use three distinct sets of data for this study, each containing individually-generated short response text on topics of varying political salience. The first dataset provides the richest bank of personality measures and is used for the majority of the analysis, though it is also a fairly small dataset with a sample of only 62 respondents. I therefore supplement my analysis with data from a nationally-representative survey about the Affordable Care Act as well as data from a sample of Twitter users who have engaged in conversations about President Trump.

The first set of data comes from a study I conducted on Mechanical Turk in Spring 2017. In this study, subjects were presented with this question taken from the General Social Survey: "Do you think abortions should be legal under any circumstances, legal only under certain circumstances, or never legal under any circumstances?" (Smith et al., 2012). Respondents were prompted to answer this question and explain their reasoning and were required to provide at least 100 words in response. After submitting their free response text, users completed a battery of survey measures, including basic demographic information, personality measures from Moral Foundations theory (Haidt and Joseph, 2008) and the Big 5 (John and Srivastava, 1999); measures of deliberativeness (Gastil et al., 2012); political knowledge (Carpini and Keeter, 1993); political ideology (Center, 2017) as well as Likert-scale questions about their position on abortion. Respondents were paid about $7 and took around 30 minutes to complete the essay and survey. The short essays submitted through this survey averaged around 120 words and were of quite high quality.

While this dataset only has 62 respondents, the broad battery of psychometric measures makes it a promising tool for benchmarking the usefulness of the conceptual network measure. If these measures of connectivity, complexity, and hierarchy truly capture something about an individual's reasoning structure we would expect them to be highly correlated with known personality metrics.

Specifically, we would expect that people who tend to have connected networks would have more political knowledge (thus seeing more interconnections) and be more conscientiousness (thus making an effort to connect every point). We might similarly expect these people to score higher on neuroticism. In line with the particularist view, we would expect subjects with more complex networks to have tendency for openness while, in accordance with the utilitarian approach, subjects with hierarchical networks my have a tendency for authority.

We may also expect to see correlations between ideology – the strongest predictor of one's position on a political issue – and the ways in which one structures their reasoning. Here, though, is where it becomes particularly interesting to separate the *content* of reasoning from the *structure* of reasoning. We would certainly expect the content of one's position to be substantially correlated with ideology, but the same is not necessarily true of structure. We would expect to see such correlations, however, if we follow Haidt (2012) in thinking that latent personality traits are the core drivers of ideology. Under this conception, liberals follow the particularist model and are primarily concerned with minimizing harm and maximizing fairness, while conservatives follow the utilitarian approach and tend towards authority and centrally guiding principles. We thus might expect complex network structure to be positively correlated with ideology while hierarchical network structure is negatively correlated with ideology.

This, though, would be a poor sign for democracy. If liberals and conservatives not only disagree on their positions, but fundamentally differ in how they think about issues and justify their reasoning, it would suggest that polarization is just a side effect of the human condition and that no amount of discourse can lead to collaborative government. If, on the other hand, ideology is not correlated with reasoning structure, it would provide some hope that, while people may not always agree, they can at least find ways to productively discuss these issues and reason together.

Since ideology is a dimension of particular interest, I also apply this method to two other datasets for which user ideology is known or can be inferred. These datasets are both larger than the MTurk dataset, but unfortunately do not include the full battery of personality measures.

The first of these datasets is a nationally representative survey conducted by the Kaiser Family Foundation to gauge opinions about the Affordable Health Care Act (AHCA). Conducted by phone in July 2017, the poll elicited respondent favorability to AHCA and then asked, "Could you tell me in your own words what is the main reason you have (a favorable/unfavorable) opinion of this proposed new health care plan?" Transcribed responses are included for 1018 subjects, with only 160 declining to respond to this prompt. This text tends to be much shorter than the responses elicited through Mechanical Turk, averaging only about 14 words However, the survey also includes a measure of ideology and political knowledge.

Finally, since my primary interest in inferring individual conceptual networks from text is as a tool for supporting political conversation, I turn to a setting where political conversation is often studied – Twitter. Here, I take a corpus of conversations from Fall 2018 which mention President Trump and examine the tweets of 520 highly active users Drawing upon the well-noted fact that Twitter user ideology can be readily discerned from who they choose to follow (Barberá, 2015), I collect each user's following list and scale user ideology as the proportion of right- versus left-leaning media accounts

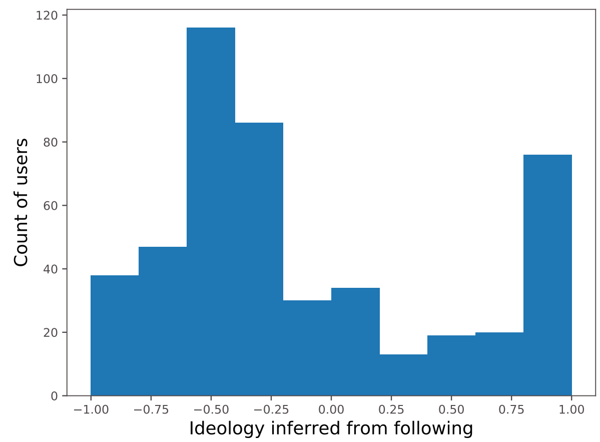


Figure 1: Ideology inferred from accounts a user is following

found. My list of well-known media accounts with well-identified ideologies includes pundits[1], media outlets[2], and politicians[3]. All benchmark accounts are normalized to the same three-point scale: -1 indicates a left left-leaning account, 0 indicates a neutral account, and 1 indicates a right-leaning account. In total, I collected a list of over 1000 benchmark accounts which are equally distributed across the three ideological categories.

Figure 1 shows the distribution of ideology scores inferred from the accounts followed by the 520 users in our sample. This distribution is consistent with what we would expect from a corpus built around conversations mentioning President Trump. On the right, the greatest mass is at the extreme - indicating die-hard conservatives who are outspoken in defending the president. On the left, we see a different pattern: users who are more moderate as well as users who are extremely liberal appear less frequently in the corpus. This suggests that moderates may have mixed-sympathies for both sides of an argument, may simply find the dialogue to be unproductive and vitriolic, or may be less prevalent in the population overall. Those on the extreme left may similarly feel that it is not worth their time or energy to argue on these issues. Users who are solidly left-leaning but not extreme, however, appear most frequently in the corpus. This suggests a contingent of users who either enjoy arguing or remain committed to dialogue across ideological divides; who find these conversations valuable enough to continue engaging.

[1] https://www.dataforprogress.org/blog/2018/11/19/identifying-and-estimating-the-ideologies-of-twitter-pundits
[2] https://www.adfontesmedia.com
[3] https://ballotpedia.org/List_of_current_members_of_the_U.S._Congress

While inferring ideology on Twitter takes some work, it is unfortunately not possible to construct a valid measure of political knowledge using this corpus. I include this measure as a placeholder in Figure 3 in order to make comparing across datasets easier.

Additionally, because this corpus is constructed particularly around conversations, I have multiple tweets from each user, and focus here particularly on users who are active in the conversation. On average, I have 25 tweets per user in my sample, with an average of about 600 words total across their tweets.

Finally, it may be that the coarse structural features identified by this method could be more simply captured using existing text analysis methods. For each dataset I therefore also measure the sentiment (compound VADER) and Flesch-Kincaid scores of all responses.

6 Results

I begin here by examining the dataset collected through Mechanical Turk, which captures 62 individuals' reasoning on the topic of abortion. Significant correlations between the battery of personality measures and three dimensions of structural reasoning are shown in Figure 2.

While we see many weak correlations between personality measures and the dimensions of connectivity, complexity, and hierarchy, many of these correlations are not what we would expect. Most notably some of the strongest positive correlations occur between the Moral Foundations dimension of authority and the structural dimension of connectivity as well as between the Moral Foundations dimension of Ingroup preference and the structural dimension of complexity. The latter, in particular seems to go against the expectations of Haidt and Joseph (2008), which would suggest that liberals have lower Ingroup preference while having more complex structure in their reasoning.

There are, however, several correlations which fall in line with our hypotheses. There does appear to be a positive correlation between conscientiousness and connectivity, as well as between preference for authority and hierarchical network structure. It is, of course, difficult to draw conclusions from this single dataset.

Finally, while there may be some minor, negative correlations between sentiment, Flesch-Kincaid and the structural dimensions, it does appear from this dataset that these measures of connectivity, complexity, and hierarchy may be picking up something which is not captured by existing textual measures.

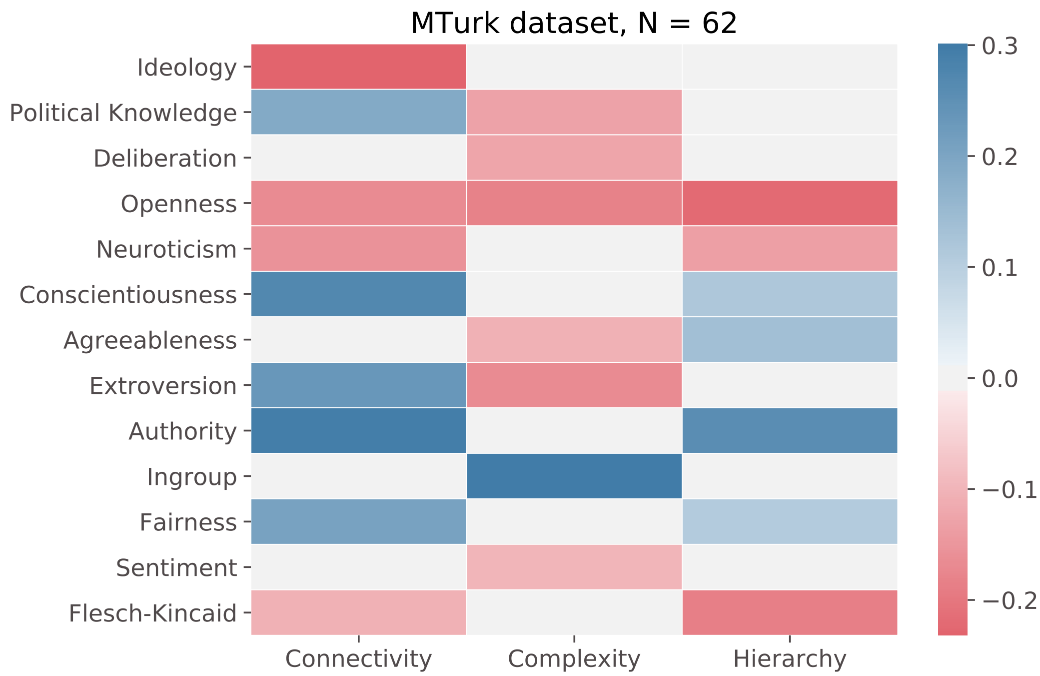


Figure 2: Correlations between personality traits and reasoning structure ($p < 0.05$)

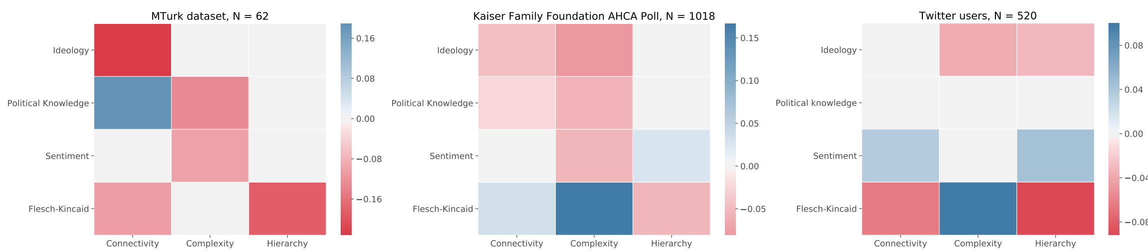
While the datasets from the Kaiser Family Foundation and Twitter do not allow for measuring detailed personality traits, I am able to capture political ideology, sentiment, and Flesch-Kincaid scores for all datasets as well as political knowledge for the Kaiser Family Foundation dataset. While these datasets each have their own strengths and weaknesses, they serve as a helpful point of reference in determing whether this method can be applied broadly to individual response text and whether there is meaning in doing so.

Figure 3 shows correlations for comparable data across all three datasets. Note that political knowledge is not available in the Twitter dataset and that this trait is measured differently between the Mechanical Turk and Kaiser Family Foundation datasets.

Here we see that all three datasets pick up on weak, negative correlations between ideology and at lease one of the structural dimensions. While this is not enough to draw any definitive conclusions, it does suggest that, unlike content, the structure of reasoning may differ from ideological stance.

In our two larger datasets, the Kaiser Family Foundation and Twitter, we see moderate correlations between Flesch-Kincaid score and complexity. While this makes sense in terms of what each measure aims to capture, id does raise the question of whether a network approach is needed in order to measure the complexity of reasoning structure.

Finally, while none of the correlations found are particularly strong, when taken together these results do suggest that the structural dimensions of reasoning are picking up on something and are worth exploring further, particularly in the context of conversational engagement.



(a) MTurk (b) Kaiser Family Foundation (c) Twitter

Figure 3: Correlations between personality traits and reasoning structure across datasets ($p < 0.05$)

7 Conclusion

In this paper, I presented a new framework for understanding individual-level reasoning structure, as separate from topical content, and presented a new method for inferring that structure from text. I applied this method to data from a Mechanical Turk study I conducted, a nationally-representative telephone poll, and a sample of politically engaged Twitter users.

This work suggests promising new ways to interpret differences in ideology and moral orientations. Moral foundations theory (Haidt, 2012) argues that political opinions are driven by an individual's orientation along at least five moral dimensions. Political divides, then can be traced back to fundamental differences in the weighting of these moral dimensions. Specifically, the moral foundations composite score for progressivism considers aversion for harm and concerns about fairness to be quintessentially liberal traits, while the dimensions related to in-group loyalty, respect for authority, and purity are more common among conservatives. Thus we would expect an individual's moral foundations progressivism score (Haidt, 2012) to be correlate with their political ideology score (Center, 2017). For both these measures, we would expect conservative thought to be more likely to result in hierarchical, hub-and-spoke like networks, where concepts differ significantly in importance (as measured by degree) and high-degree nodes tend to connect to low-degree nodes. Such networks are representative of the utilitarian view (Sidgwick, 1907), where a few core rules dictate judgments. Progressive networks, on the other hand, would tend to be more decentralized, without a core driving principal, in line with particularist moral theories (Dancy, 1993).

While this study finds no evidence that liberals are more likely to have complex networks or that conservatives are more likely to have hierarchical networks, it does demonstrate that reasoning structure can be inferred from short text and suggest that there may be behavioral meeting to this structure beyond ideology.

In future work, I plan to apply this method to additional datasets, particularly surveys with rich batteries of personality measures. Finally, I plan to use this method as a tool for evaluating the productiveness of political conversations, e.g., as a measure of whether people are talking past each other or actually engaging with each other.

References

Abbott, R., Walker, M., Anand, P., Fox Tree, J. E., Bowmani, R., and King, J. (2011). How can you say such things?!?: Recognizing disagreement in informal political argument. In *Proceedings of the Workshop on Languages in Social Media*, pages 2–11. Association for Computational Linguistics.

Atteveldt, W. v. (2017). Crowd sourcing for sentiment analysis. In *International Communications Association*.

Austin, J. (1962). *How to Do Things With Words (The William James Lectures)*. Harvard University Press.

Axelrod, R. (1976). *Structure of decision: The cognitive maps of political elites*. Princeton university press.

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis*, 23(1):76–91.

Berker, S. (2015). Coherentism via graphs. *Philosophical Issues*, 25(1):322–352.

Blum, R. (2019). *Party Takeover: the Tea Party and Party Factions in the U.S.* University of Chicago Press.

Campbell, Angus; Converse, P. E. M. W. E. S. D. E. (1960). *The American Voter*. University of Chicago Press.

Carley, K. (1993). Coding choices for textual analysis: A comparison of content analysis and map analysis. *Sociological Methodology*, 23:75–126.

Carpini, M. X. D. and Keeter, S. (1993). Measuring political knowledge: Putting first things first. *American Journal of Political Science*, pages 1179–1206.

Center, P. R. (2017). Are telephone polls understating support for trump? Report.

Christen, M. and Ott, T. (2013). Quantified coherence of moral beliefs as predictive factor for moral agency. In *What Makes Us Moral? On the capacities and conditions for being moral*, pages 73–96. Springer.

Cohen, J. (1989). *Deliberation and Democratic Legitimacy.* The Good Polity: Normative Analysis of the State.

Cohen, R. (1987). Analyzing the structure of argumentative discourse. *Computational linguistics*, 13(1-2):11–24.

Collins, A. M. and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6):407.

Dancy, J. (1993). Moral reasons.

Dewey, J. (1927). *The public and its problems.* Swallow Press.

Dorsey, D. (2006). A coherence theory of truth in ethics. *Philosophical studies*, 127(3):493–523.

Dorsey, D. W., Campbell, G. E., Foster, L. L., and Miles, D. E. (1999). Assessing knowledge structures: Relations with experience and posttraining performance. *Human Performance*, 12(1):31–57.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in cognitive sciences*, 7(10):454–459.

Eveland, W. P. and Cortese, J. (2004). How web site organization influences free recall, factual knowledge, and knowledge structure density. *Human Communication Research*, 30(2):208–233.

Feng, V. W. and Hirst, G. (2011). Classifying arguments by scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 987–996, Portland, Oregon, USA. Association for Computational Linguistics.

Gastil, J., Knobloch, K., and Kelly, M. (2012). *Evaluating Deliberative Public Events and Projects*, book section 10. Oxford University Press, Oxford; New York.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., and Ditto, P. H. (2011). Mapping the moral domain. *Journal of personality and social psychology*, 101(2):366.

Habermas, J. (1984). *The theory of communicative action.* Beacon Press, Boston.

Habernal, I. and Gurevych, I. (2015). Exploiting debate portals for semi-supervised argumentation mining in user-generated web discourse. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2127–2137.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage.

Haidt, J. and Joseph, C. (2008). *The Moral Mind: How Five Sets of Innate Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules*, volume 3. Oxford University Press.

Haste, H. (2013). Deconstructing the elephant and the flag in the lavatory: Promises and problems of moral foundations research. *Journal of Moral Education*, 42(3):316–329.

Hong, L., Page, S. E., and Baumol, W. J. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 101, pages 16385–16389.

John, O. P. and Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999):102–138.

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

Katzav, J. and Reed, C. (2008). Modelling argument recognition and reconstruction. *Journal of Pragmatics*, 40(1):155–172.

Lazarsfeld, PF; Berelson, B. G. H. (1948). *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. New York, Columbia University Press.

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

Mansbridge, J. (2015). A minimalist definition of deliberation. *Deliberation and development: Rethinking the role of voice and collective action in unequal societies*, pages 27–50.

McNaughton, D. and Rawling, P. (2000). Unprincipled ethics. *Hooker and Little*, 2000:256–275.

Mochales, R. and Moens, M.-F. (2011). Argumentation mining. *Artificial Intelligence and Law*, 19(1):1–22.

Navigli, R. and Ponzetto, S. P. (2012). Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Newman, M. E. (2003). Mixing patterns in networks. *Physical Review E*, 67(2):026126.

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Systems Research and Behavioral Science*, 12(5):410–430.

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.

Rawls, J. (1993). *Political Liberalism*. John Dewey essays in philosophy. Columbia University Press.

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A. A., and Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning. *International Journal of Learning and Media*.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423.

Shapiro, S. C. and Rapaport, W. J. (1987). Sneps considered as a fully intensional propositional semantic network. In *The knowledge frontier*, pages 262–315. Springer.

Shavelson, R. J. (1974). Methods for examining representations of a subject-matter structure in a student's memory. *Journal of Research in Science Teaching*, 11(3):231–249.

Sidgwick, H. (1907). *The methods of ethics*. Hackett Publishing.

Smith, T. W., Marsden, P., Hout, M., Kim, J., and Davis, J. (2012). General social surveys. *National Opinion Research Center*.

Speer, R. and Havasi, C. (2012). Representing general relational knowledge in conceptnet 5. In *LREC*, pages 3679–3686.

Stab, C. and Gurevych, I. (2014). Identifying argumentative discourse structures in persuasive essays. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 46–56, Doha, Qatar. Association for Computational Linguistics.

Thagard, P. (1998). Ethical coherence. *Philosophical Psychology*, 11(4):405–422.

Toulmin, S. E. (1958). *The uses of argument*. Cambridge University Press.

V. O. Key, J. (1942). *Politics, parties and pressure groups*. New York, Thomas Y. Crowell Company.

Walton, D. N. (1996). *Argumentation Schemes for Presumptive Reasoning*. L. Erlbaum Associates.

Zaller, J. R. et al. (1992). *The nature and origins of mass opinion*. Cambridge university press.