Opinion Dynamics Model Based on Cognitive Biases of Complex Agents

Pawel Sobkowicz¹

¹KEN 94/140, 02-777 Warsaw, Poland

Correspondence should be addressed to pawelsobko@gmail.com

Journal of Artificial Societies and Social Simulation 21(4) 8, 2018

Doi: 10.18564/jasss.3867 Url: <http://jasss.soc.surrey.ac.uk/21/4/8.html>

Received: 20-01-2018

Accepted: 17-09-2018

Published: 31-10-2018

Abstract: We present an introduction to a novel way of simulating individual and group opinion dynamics, taking into account how various sources of information are filtered due to cognitive biases. The agent-based model presented here falls into the ‘complex agent’ category, in which the agents are described in considerably greater detail than in the simplest ‘spinon’ model. To describe agents’ information processing, we introduced mechanisms of updating individual belief distributions, relying on information processing. The open nature of this proposed model allows us to study the effects of various static and time-dependent biases and information filters. In particular, the paper compares the effects of two important psychological mechanisms: confirmation bias and politically motivated reasoning. This comparison has been prompted by recent experimental psychology work by Dan Kahan. Depending on the effectiveness of information filtering (agent bias), agents confronted with an objective information source can either reach a consensus based on truth, or remain divided despite the evidence. In general, this model might provide understanding into increasingly polarized modern societies, especially as it allows us to mix different types of filters: e.g., psychological, social, and algorithmic.

Keywords: Opinion Change, Motivated Reasoning, Confirmation Bias, Complex Agents, Agent-Based Model

Introduction

- 1.1 The actual processes through which individual people and groups of people evaluate information and form or change their opinions are very complex. Psychology offers many descriptions of these processes, often including multiple pre-conditions and influencing factors. The assumption that opinions form through truth-seeking, rational reasoning is unfortunately not true in most cases. The list of recognized cognitive biases that influence our mental processes (rational and emotional) is very long, covering over 175 named entries (Benson 2016). The situation becomes even more complex when we try to describe how changes of individual opinion combine to form dynamical social systems. In addition to the problems alluded to above, one has to consider the multiple forms of social interactions: personal (fact to face and especially in recent years, those mediated by electronic media) and public (news, comments, rumours and other modes of information reaching an individual). These interactions vary in their informative and emotional content, trust of source of information, its pervasiveness and strength to name a few. Taking these difficulties into account, the task of an accurate description of individual and group opinion change dynamics appears insurmountable. Yet, the need to understand how and why our societies (especially democratic ones) arrive at certain decisions, how and why people change their beliefs (or why they remain unconvinced in the light of ‘overwhelming evidence’), what are the mechanisms driving the increasing polarization of our societies and how to make people talk to and understand each other, is so great that despite the challenges, there is considerable research on this topic.
- 1.2 One of the most active discussions in psychology of belief dynamics, is centred around apparently irrational processing of information. This covers operations of biases, heuristic short-cuts and other effects, which compare with the classical tenets of rational choice theory. Apparently **irrational** behaviour has not received much attention within the ABM community so far, despite its presence in many social situations. Important examples are provided by strong opposition to well documented arguments in cases of climate change, vaccination, energy policies etc. There are well known differences in risk perception and reactions, leading to strong polarization almost beyond the capacity to communicate (Tversky & Kahneman (1974, 1981, 1986); Opaluch & Segerson (1989); Tversky et al. (1990); Kahneman (2011); Sunstein (2000, 2002, 2006); Sunstein et al. (2016)).

- 1.3 Biases in information processing are also recognized within approaches of **rational** behaviour, for example in the Subjective Expected Utility model (Sugden 2003). Differences in goals (foreground and background), their changes in time may influence how information is processed and decisions are made. For a review of such rational, framing-based origins of biases see Lindenberg (2001, 2010).
- 1.4 Our current work has been motivated by recent studies (Kahan (2016a,b)), which describe in detail the Politically Motivated Reasoning Paradigm (PMRP). These results, which have shown that it might be possible to differentiate findings due to rather subtle differences between potentially applicable psychological biases (confirmation bias and motivated reasoning, described in detail in Sections 1.10-1.13).
- 1.5 For several years, group opinion change has been a fertile ground for socio-physics and Agent Based Modelling (ABM). Initial research has used many of the tools and ideas developed to describe magnetic phenomena and analogies between atomic spin states and opinions, magnetic field and external influences to derive statistical descriptions of global opinion changes. There are many approaches, for example the voter model (Cox & Griffeath 1986; Ben-Naim et al. 1996; Galam et al. 2002; Castellano et al. 2003), the Sznajd model (Sznajd-Weron & Sznajd 2000; Stauffer 2001; Stauffer & de Oliveira 2002; Stauffer 2002; Slanina & Lavicka 2003; Sabatelli & Richmond 2004, 2003; Bernardes et al. 2001), the bounded confidence model Deffuant et al. (2000, 2002); Weisbuch (2004); Weisbuch et al. (2003), the Hegselmann-Krause model (Hegselmann & Krause 2002), the social impact model of Nowak-Latané (Nowak et al. 1990; Nowak & Lewenstein 1996) and its further modifications including the role of leaders (Holyst et al. 2001; Kacperski & Holyst 2000, 1999; Sobkowicz 2010), and many more others. Historically, the initial focus was on the formation of consensus – treated as a form of a phase transition – but later work focused on the role of minorities, with special attention given to the effects of inflexible, extremist individuals.
- 1.6 The literature on numerical models of opinion dynamics has grown enormously over the past decade. For relatively recent reviews we can point out Castellano et al. (2009); Castellano (2012); Galam (2012). While most early work was limited to studies of the **models themselves** (rather than specific social contexts), with very interesting sociophysical results, but only weak, qualitative correspondence to any real societies (Sobkowicz (2009)), recently this situation has changed. The availability of large scale datasets, documenting opinions and interactions between people (derived mainly from the Internet and social media), has allowed us in principle, to attempt quantitative descriptions of specific opinion evolution processes. The number of socio-physical and ABM based work aimed at quantitative description of real societies remains however limited. For example, in the case of political elections, only a few papers have attempted such descriptions (Caruso & Castorina (2005); Fortunato & Castellano (2007); Palombi & Toti (2015); Sobkowicz (2016); Galam (2017)).
- 1.7 Despite these undoubted advances, **the socio-physical models of individual behaviour are still rather crude**, even in their fundamental assumptions. For example, most socio-physical descriptions of agent's individual behaviour are too simplistic, too 'spin-like', and thus unable to capture the intricacies of human behavioural complexity. This observation also applies to descriptions of interactions between agents, or more generally way, to new information is treated in the process of adjusting currently held opinions. Similarly, many Agent Based Models assume relatively simple forms of such interactions, for example rules which state that if agents are surrounded by other agents holding different opinion to their own, they would change their opinion to conform with the majority. While this assumption is validated by the conformity experiments of Asch & Guetzkow (1951); Asch (1955) and their follow-ups (Shiller 1995; Bond & Smith 1996), the willingness to radically change an important belief (political, religious, social) is much smaller. As experience with real life situation shows, such 'forced' conversion is rather unlikely among people (in contrast with atomic spins...), especially related to issues of high emotional commitment. The difference between model behaviour of spin-persons (*spinsons*, Nyczka & Sznajd-Weron (2013)) and psychology based understanding of real people has forced the introduction of special classes of agents, behaving in a way that is different from the rest: conformists, anti-conformists, contrarians, inflexibles, fanatics... Using appropriate mixtures of 'normal' and special agents it has been possible to make the models reproduce more complex types of social behaviour.
- 1.8 Such artificial division of agents into separate classes with different but fixed internal dynamics, while improving the models' range of results, seems psychologically incorrect. Anyone may behave inflexibly or show contrary behaviour in a specific situation, given the right encouragement, priming or with an appropriate framing of the issue in question. There are 'complex agent' models, in which opinion change results from a combination of agents' information and emotional state, coupled with the informative and emotional content of the message processed by the agent (which may originate from an interaction with another agent or from the media). For example, a model that non-linearly links incoming information and emotions (Sobkowicz 2012, 2013a), has given us a quantitative description of an Internet discussion forum (Sobkowicz 2013b) and even predictions of recent elections in Poland (Sobkowicz 2016). The model applies however, only to situations in which the emotional component is very strong, dominating individual behaviour.

1.9 The subtlety of individual human behaviour is typically absent in current opinion-dynamics literature. Of course, due to the fast growth of the opinion modelling field, we may have missed some important contributions which use internally complex and flexible agents. To a large extent however, it seems that most papers are based on relatively simple internal mechanisms for agents (update rules). The complexity and apparent irrationality of how real people change their opinions, such as obstinacy, contrariness, backfire effects, increasing polarization, to name but a few, are addressed by *ad hoc* modifications. The current contribution purposely ignores most important “social” aspects of traditional models: i.e., interactions between agents, social network topology and dynamics etc. Rather, we have introduced more complex internal mechanisms of information processing. More complex and bias-dependent – yet hopefully, still simple enough to be viable for modelling. The inclusion of social networks and interactions is an obvious next step to achieve realism in descriptions of actual social situations, such as achieving a consensus in a group (or lack of it). The scope of this paper is limited on purpose, for two reasons:

- to propose and describe a universal framework for complex agents influenced by numerous biases;
- to check whether the framework, applied to an extremely simple situation, leads to intuitively sensible results.

The application to more complex systems, including social interactions is left for future work.

1.10 Dissatisfaction with the spin-analogy and other simplistic models, due to inadequacy in describing the foundations of individual opinion change mechanisms has prompted some researchers to move to more psychology-based models. Among these, a small but important part is based on the Bayesian framework. Despite the recognized status of Bayesian updating in risk assessment and other areas, it is seldom used by the ABM community. In a series of papers using this approach, Martins explored the approach, starting with the Continuous Opinion Discrete Action (CODA) model (Martins 2008, 2009), continuing in Martins & Kuba (2010), and finally analysing certain spin-based, discrete opinion models as limiting cases of the CODA model (Martins 2014). In the original CODA model (Martins 2008), if an agent observes another agent acting as if they believed in an opinions, the observer should change his/her own belief to more be favourable for this opinion. The only information is through observed actions, not internal beliefs. The expanded version (Martins 2009), uses continuous variables for opinions, described not only by average values but also by uncertainty, using normal distribution of likelihood. This paper is mentioned here specifically because it is the closest in approach to our framework – and because in principle, it could also be used to describe Kahan’s empirical studies.

1.11 There is more Bayesian opinion dynamics research. To mention a few examples, Suen (2004) considered the effects of information coarsening (due to agents’ reliance on specialists for relevant information) and the tendency to choose the sources which confirm pre-existing beliefs; Bullock (2009) studied the conditions in which peoples’ beliefs, updated using Bayesian rules could in the short term, rather of converge on a true value, diverge or even polarize. Ngampruetikorn & Stephens (2016) analysed the role of confirmation bias in consensus formation in a binary opinion model on a dynamically evolving network.

1.12 The Bayesian approach allows much greater complexity of the behaviour of individual agents and offers potentially more relevant descriptions of social behaviour than spin-based models. Of course, these benefits do not come without a price; there are many more degrees of freedom in the system and therefore many more unknowns quantities in properly setting up ABM simulations. Still, the importance of social phenomena observed around the world, in particular various forms and effects of polarization, suggests the need for a more thorough understanding of its underlying mechanisms, and makes the effort worthwhile.

1.13 **We should note, however, that our framework is not Bayesian in a strict sense.** It uses similar technical rules, but filtering is not provided purely by prior beliefs: it may be the result of external influences or internal biases which are fixed and unchanging. We used some Bayesian terminology for their intuitive simplicity (e.g. prior/posterior beliefs), but the model extends beyond direct Bayes reasoning. Only in case of the confirmation bias (where the current opinion of the agent serves as the filter, as will be discussed in Section 1.13) the model approaches a fully Bayesian framework (subject to modifications like the memory reset effect, Sections 2.10-2.12). In contrast to Martins (2009), the motivated reasoning in our model is not based on pairwise exchanges of information between agents, but on a comparison with the information about averaged belief for the social group that the agent in question ‘belongs to’. For motivated reasoning, this is a fundamental difference, as information about the ‘norm’ beliefs of ones’ own group may come not only from direct observations, but also from external sources: news, media. It may also be manipulated and untrue. Our approach allows several areas of flexibility of application: i.e., different information sources (which may be due to contacts with other agents or rumours or media), different sources of filters (internal and external), different types of filters (both static and evolving) and their application (Figure 1); as well as complexity of individual information processing (Figure 2).

It is our hope, that this flexibility would allow us to eventually construct quantitative descriptions of specific social situations.

Confirmation bias vs. Politically Motivated Reasoning Paradigm

- 1.14** One of the best recognized biases in information processing is *confirmation bias*, defined by Encyclopaedia Britannica as ‘the tendency to process information by looking for, or interpreting, information that is consistent with one’s existing beliefs’. Such definition stresses that the operation of confirmation bias may be on various levels: selecting and preferring the information sources, giving different weight to different sources and internal mechanisms (such as memory preferences or storing/recall of information). Confirmation bias may be regarded as ‘internal’, in the sense that it is related directly to an actual mental state of the person (preferences, beliefs, memories) – but it is only indirectly related to person’s goals and desires. When people communicate, the individual confirmation bias effects may be combined in a way that creates group effects such as echo chambers. As a result, even when faced with true information, the agent (or a group of agents) may form or maintain a false opinion due to the confirmation bias.
- 1.15** In contrast, the *motivated reasoning* paradigm considers how goals, needs and desires directly influence the information processing (Jost et al. (2013)). These goals may be related to individual needs, but also to group or global ones, for example the goal of *achieving or maintaining the person’s position within a social group*. In such cases, motivated reasoning may bias information processing by substituting the goal of truth-seeking by a person’s desires to affirm the affiliation with the chosen in-group. If these goals are connected to social politics, the process can be called Politically Motivated Reasoning (PMR) paradigm. Seen through the lens of these desires, the apparently irrational choices (such as disbelief in well documented evidence and belief in unproven claims) become rational again. As the goal of the person shifts from truth seeking to strengthening of the position within a social group, disregard for the truth becomes rational. Especially, when the consequences of rejection from the group is more immediate and important than ‘erroneous’ perceptions of the world. Kahan (2016a,b) has provided a very attractive framework, allowing us not only to describe the role of various forms of cognitive biases, but also the empirical evidence of differing predictions of different heuristics, such as confirmation bias or political predispositions. Experiments with manipulated ‘evidence’, described by Kahan, are very interesting. While both these mechanisms lead away from truth-seeking behaviour, their predictions might differ, especially with respect to new information. The confirmation bias depends on internal agent states, while PMR involves perception of external characteristics.
- 1.16** The vision of information processing, comparing these two forms of bias, described by Kahan, is simple enough to become an ABM framework. As we shall argue, the information filtering approach is very flexible and could be applied to a variety of situations, contexts and types of processing bias. **Our present goal is to describe such a framework and to provide simple examples of the types of information processing leading to consensus or polarization.** The latter is of special importance, as the current political situation in many democratic countries seems to be irrevocably polarized, with large social sections unable to find common ground on many important issues.
- 1.17** Examples of such polarization, in which certain sub-communities not only disagree, but are, in practice, incapable of communication, are quite numerous. Probably the best studied example is the increase of the split between conservatives and liberals in the US (represented politically by the Republicans and the Democrats), although it is increasingly being found in other countries and smaller communities. Examples include growing splits of opinions on specific subjects (Brexit, the immigration crisis, GMO, climate change, vaccination, to name a few), for which the upholders of specific views are largely disregarding the information used by the opponents, living in “echo chambers” or “filter bubbles”. The phenomenon is very visible in on-line environments, but also in traditional media, which increasingly serve partisan interests.

Individual Information Processing Model

- 2.1** The current work aims at a general, flexible model of the **individual** opinion dynamics. No inter-agent communication is considered in the model at this stage, but the model allows a natural extension to a socially connected situation. This is by treating the information received from other agents on similar footing as non-personal external sources. In fact, the range of biases and idiosyncrasies in such a situation could be greater, due to the rich ways people relate to each other, socially and emotionally.

2.2 We base our concepts on the quasi-Bayesian belief updating framework. Figure 1 presents the basic process flow, modelled after Kahan (2016a). For simplicity, we shall assume that the belief which we will be modelling may be described as a single, continuous variable θ , ranging from -1 to +1 (providing a natural space for opinion polarization). The agent holds a belief on the issue, described at time t by a distribution $X(\theta, t)$. For example, if the agent is absolutely sure that the ‘right’ value of θ is θ_0 , then the distribution would take the form of Dirac delta function centred at θ_0 . Less ‘certain’ agents may have a different form of $X(\theta, t)$. This distribution is taken as a prior for an update, leading to the opinion at $t + 1$. In the simplest case, the likelihood factor would be provided by the new information input $S_i(\theta)$. Here the index i corresponds to various possible information sources. Kahan proposed that rather than this direct update mechanism (prior opinion+information→posterior opinion), **the incoming information is filtered by the cognitive biases or predispositions of the agent**. The filtering function $F(S_i)$ transforms the ‘raw’ information input $S_i(\theta)$ into the filtered likelihood $FL(S_i, \theta)$, so that the posterior belief distribution $X(\theta, t + 1)$ is obtained by combining the prior opinion distribution $X(\theta, t)$ with the likelihood filter $FL(S_i, \theta)$. It is important to note that different sources of information may be filtered in different ways. Trust in the source, cognitive difficulty of processing the information, its emotional context, the agent’s dominant goals – they all may influence the ‘shape’ of the filter.

Figure 1: Basic model of information processing. An agent holds a prior belief about an issue, described by a distribution $X(\theta, t)$. We assume a simple, one-dimensional ‘opinion parameter’ θ ranging from -1 to 1. The information on the issue, coming from source S_i has a distribution $S_i(\theta)$. This information is filtered by a function $F(S_i)$, specific to the information source. The form of the filtering function may vary, depending on the specific bias considered in the model. Combining the information input with the filter function yields the filtered likelihood information $FL(S_i)$. The update of the agent’s belief $X(t)$ via $FL(S_i)$ leads to the changed, posterior distribution of beliefs $X(\theta, t + 1)$.

- 2.3** Information that influences people’s beliefs comes from multiple sources. The list below (definitely not complete) divides these into three groups: personal/individual, in-group focused and out-group (the latter including media).
- 2.4** There is, of course, **personal experience**, which may provide high impact information about specific facts and events, and, with the application of certain cognitive processes, trends, estimates, diversity and prognoses. Direct experiences may be thought of as personal and therefore trustworthy, but in many cases we rely on memory, which may provide false information. A well-known example of such effect is the availability bias, in which more weight is given to information more easily accessible. Some other cognitive biases are also relevant for personal observations; we may fall for certain illusions, disregard a part of an experience and put emphasis on other parts, even to the degree of actually inventing events that did not take place.
- 2.5** The second source of information is related to the group of people with whom a person identifies (the in-group). These inputs may come from **in-group information exchanges**, either in person or via electronic or traditional communication media. The latter has become increasingly important during the past decade, especially among the younger population. In addition to the interactions with specific individuals in the in-group, the in-group may influence agents’ beliefs via cumulative indicators. These would include the **official or semi-official statements of the group’s views on specific issues**, but also the **unofficial and media information about group norms, average opinions and trends**. The latter are especially interesting, as they may derive both from within

the group and from outside. In such cases the information about in-group views and norms may be manipulated and distorted.

- 2.6** The last group of the sources is related to any source outside the in-group. This may include **interactions with people outside one's own self-identification group** and the **media** perceived as not associated with the in-group. In case of the media the information is **prepared** by someone, which includes both the selection and presentation of the information (e.g. party political manifestos and programs).
- 2.7** The information which we use to fortify or to change our beliefs may be manipulated 'at source'. In personal interactions with other people we may get the wrong impressions due to many forms of dishonesty or distortion. Traditional sources of news are also subject to misrepresentation. The ideal of fair and balanced journalism – giving comparable attention to all contradicting views – may also, at times, be considered manipulative, especially when it results in undue attention and coverage given to a tiny minority of views. An example of negative consequences of such 'balanced' reporting may be provided by the case of the anti-vaccination movement (Betsch & Sachse (2013); Nelson (2010); Tafuri et al. (2014); Wolfe et al. (2002)).
- 2.8** In reality, however, manipulations due to **unbalanced reporting** are much more frequent, e.g. the partisan media phenomenon. The polarization of both traditional channels (newspapers, radio, TV) and Internet sources (WEB versions of the traditional channels and independent WEB pages, blogs, Facebook pages and tweets) is a well-known phenomenon (Adamic & Glance (2005); Stroud (2010); Campante & Hojman (2010); Lawrence et al. (2010); Jerit & Barabas (2012); Prior (2013); Wojcieszak et al. (2016)). Many people rely on a limited number of information sources, the spectrum of information reaching him/her could be heavily distorted. This selective attention/selective exposure can lead to the echo-chamber phenomenon, where a person sees and hears only the information supporting the 'right' beliefs.
- 2.9** **In the current work we shall use an arbitrarily chosen form of the source information**, $S_T(\theta)$, assumed to be a Gaussian distribution centred around the 'true' value θ_T . This particular choice of the source information form is motivated by the desire to study how the agents with differing initial belief distributions react to such 'approximately true' information. Are they capable of 'correcting' their views?
- 2.10** The Bayesian-like form of the filtering process, multiplying the incoming information $S_T(\theta)$ by the filter function $F(\theta)$ is very efficient: a single information processing event may decisively change the shape of the information distribution if the filter is narrow enough. For this reason, we introduced here a process-control parameter, the **filtering efficiency** f . Its role is to determine the relative strength of the influence of the specific filtering function on the incoming information. In particular, the **effective filter function** is assumed to take the form $fF(\theta) + (1 - f)U$, where U is a uniform function.
- 2.11** The filter function $F(\theta)$ may depend on multiple agent characteristics. Consider the confirmation bias as an example: $F(\theta)$ would depend not only on the current belief of the agent $X_j(\theta, t)$, but also on the importance attached to the issue in question (which can be partially described by changing the filtering efficiency f).
- 2.12** To summarize, when agents modify their opinion distribution (probability p), the equation governing the transition is

$$X_j(\theta, t + 1) = [X_j(\theta, t)FL(S_i, \theta)]_N, \quad (1)$$

where the symbol $[\]_N$ denotes normalization over the allowed interval $[0,1]$. In the three cases considered in this paper and the source $S_T(\theta)$, the filtered likelihood functions can be specified as follows. For no filtering

$$FL(S_T, \theta) = S_T(\theta). \quad (2)$$

For the case of the individual confirmation bias

$$FL(S_T, \theta) = [S_T(\theta) (fX_j(\theta, t) + (1 - f)U)]_N; \quad (3)$$

and for the case of politically motivated reasoning (PMR)

$$FL(S_T, \theta) = [S_T(\theta) (fX_G(\theta, t) + (1 - f)U)]_N, \quad (4)$$

where $X_G(\theta, t)$ is the group average of opinion distribution for the agent's in-group at time t .

- 2.13** In addition to the **strength** of a particular filter (f) in processing of information, which corresponds to the individual importance of the issue in question, we note that not all our interactions with other people or with the media are necessarily transformative. To account for this Martins (2009) has proposed a modification of the original Bayesian rules, in which only the update of the belief estimate due to interaction between two agents is weighted by a certain function, depending on the difference of opinions between two agents. If the difference is large enough, the agents influence each other only very slightly.

- 2.14** In our approach, only a fraction p of encounters with the information sources leads to the informative processing, characterized by the likelihood function $FL(S_i)$. In the remaining $1 - p$ cases, the encounter is **ignored and the information is not processed**. The simplest approach to describe such situation would be to leave the belief distribution unchanged, $X_j(\theta, t) = X_j(\theta, t + 1)$, which may be treated as in the case of the agent's **perfect memory**. However, as we shall see, repeated applications of Bayesian updates leads to a narrowing of an agent's individual belief distributions. Eventually, the beliefs would become more and more focused, influencing the dynamics of the whole system. For this reason we will introduce an **imperfect memory mechanism** that restores some level of an **individual belief indeterminacy**, in which the agent reverts partially to the intrinsic value of the standard deviation of its $X(\theta, t)$ distribution. The origin of this reset of indeterminacy can be explained by numerous encounters with a range of beliefs, other than the main source considered in simulations, which are too weak to significantly shift the agent's average opinion, but introduce some degree of uncertainty, broadening the belief distribution function $X_j(\theta)$.
- 2.15** This is described as follows: when ignoring the information event (probability $1 - p$) the agent's belief distribution does not remain unchanged but becomes

$$X_j(\theta, t + 1) = mX_j(\theta, t) + (1 - m)N(\langle\theta\rangle_j(t), \sigma_{0j}), \quad (5)$$

where the **memory fidelity parameter** $0 \leq m \leq 1$ describes the ratio of preserving the current distribution intact, and $N(\langle\theta\rangle_j(t), \sigma_{0j}(t))$ is a Gaussian distribution centred at the current average belief of the agent $\langle\theta\rangle_j(t)$, but characterized by an initial fixed standard deviation σ_{0j} , characteristic for each agent. Thus, for perfect memory ($m = 1$) distribution remains unchanged, and for $m = 0$, an agent 'left to itself' preserves the current average value of the belief, but resets the indeterminacy of its beliefs to its initial value. Information processing is graphically presented in Figure 2.

Figure 2: Details of information processing. The likelihood function $FL(S_i(\theta))$, derived from information source S_i is applied only in case of 'informative' encounter, with probability p . In such case $X_j(\theta, t + 1) = X_j(\theta, t)FL(S_i(\theta))$. In the remaining cases, the encounter is ignored. Without processing new information the agent's belief function may remain unchanged, or it may become somewhat relaxed, by addition of a Gaussian function $N(\langle\theta\rangle_j(t), \sigma_{0j})$, centred at the **current** average $\theta(t)$ for the agent, and characterized by the standard deviation equal to the **starting** value σ_{0j} . Depending on the value of the memory parameter m the posterior belief of the agent becomes then $X_j(\theta, t + 1) = mX_j(\theta, t) + (1 - m)N(\langle\theta\rangle_j(t), \sigma_{0j})$.

Basic Simulation Assumptions

- 3.1** In real situations, both the information sources and the filters described in the previous section combine their effects in quite complex ways. We encounter, in no particular order, information sources of various type, content and strength, in some cases acting alone, in others - combined. To elucidate the model's effects, we shall initially focus on drastically simplified systems, in which we show the effects of the **repeated application of the same filter to the same information source distribution** $S_i(\theta)$, for a range of starting belief distributions $X_j(\theta, 0)$ (where the index j denotes individual agents). The aim of this exercise is to show whether particular filters (no filter, confirmation bias, PMR) lead to stable belief distributions, polarization, emotional involvement etc.

- 3.2** As noted, for the simulations shown here **we shall be using the truth-related form of the information source**, $S_T(\theta)$, assumed to take a rather broad Gaussian form, centred at $\theta_T = 0.6$ and with standard deviation equal to 0.4. This choice of the information source distribution is motivated by two reasons. The first is to check whether the simulated society is capable of reaching consensus when the information source points to a well-defined value. The second reason was to study the effects of asymmetry. Obviously, it is much easier for agents whose initial opinion distribution favours positive θ values to ‘agree’ with an information source favouring a positive θ_T . In contrast, agents starting with belief distributions preferring negative θ values, would have to ‘learn’ and to overcome their initial disagreement and to significantly change their beliefs.
- 3.3** Each agent is initially characterized by a belief function of a Gaussian form (bounded between -1 and $+1$ and suitably normalized). The standard deviation parameters for agents σ_{0j} are drawn from a uniform random distribution limited between 0.05 and 0.2.
- 3.4** **Three separate sets of agents are created and used in the simulations: leftists, centrists and rightists** (we note here that these names have no connection with real world political stances and **refer to the position on the abstract θ axis**). Each agent community is composed of N agents (in the simulations we use $N = 1000$). The leftists have their initial Gaussian centre values $\langle \theta \rangle_j(0)$ drawn from a uniform random distribution bounded between -1 and -0.5 . The centrist group is formed by agents with $\langle \theta \rangle_j(0)$ drawn from between -0.5 and 0.5 , and the rightists have $\langle \theta \rangle_j(0)$ drawn from between 0.5 and 1 . The simulations discrete time steps. Time is measured in time units in which each agent in the current group has had a single chance to interact with the information source or to ignore it.
- 3.5** The program source file and datasets are available in the Open Science Framework repository at the following address: <https://osf.io/7xasr/>.

Model Results

Case 1: Unfiltered effects of true information

- 4.1** We shall start the description of the model results with a relatively simple case, with the aim of showing the effects of certain simulation parameters. The first case is based on **unfiltered** processing of the ‘truth-related’ information, $S_T(\theta)$, as shown in Figure 3. As the θ_T value is positive (equal to 0.6), the most interesting question is how such information would influence the agents who initially hold opposite views (the ‘leftists’).
- 4.2** The speed with which the average values of individual agents’ opinion distributions (their ‘preferred’ opinions) converge at the true value consensus depends on the significant information processing probability p . Figure 3 presents the time evolution of individual agents’ average beliefs ($\langle \theta \rangle_j$, thin lines) and the ensemble averages $\langle \theta \rangle_G$ for the three agent groups. We start with agents characterized by perfect memory ($m = 1$). The time evolution of the average $\langle \theta \rangle_j$ for $p < 1$ looks qualitatively different than in the case of $p = 1$. They exhibit a step-like structure, due to ‘freezing’ of beliefs if no processing takes place. However, the ensemble averages are quite similar for $p < 1$ and $p = 1$. Figure 4 shows the dependence of the time evolution of the average belief for the whole leftist group $\langle \theta \rangle_L(t)$ on the value of the parameter p (note the logarithmic scale of the time axis). In fact, a simple rescaling of the time axis to $t' = pt$ (shown in the inset) shows that the evolution is really a simple slowdown due to inactivity periods, when no information is processed. Thus, for the perfect memory (i.e., for $m = 1$), the role of p is rather trivial. It becomes more important when ‘idle’ times are used to partially reset individual uncertainty.
- 4.3** The truth-focused information flow eventually convinces all agents to believe in the ‘true’ value of $\theta_T = 0.6$ (the centre of the information source), regardless of their initial positions. The process is the fastest for agents with relatively broad-minded beliefs (high σ_{0j}). For agents with initial very narrow belief distributions the transition is shifted to later but then it is almost instantaneous (typical for Bayesian update of single valued probabilities rather than distributions). Changes in the form of belief distribution consist of a more or less gradual ‘flow’ of beliefs from the original form to a belief centred around the maximum of $S_T(\theta)$.
- 4.4** To understand the effects of the memory parameter m , it is illustrative to study the effects of the indeterminacy reset on the evolution of the **individual** opinion distributions $X_j(\theta, t)$. The relaxation of indeterminacy introduced by imperfect memory factor $m < 1$ leads to a qualitatively different final form of the individual belief distributions. Instead of a set of narrow, delta-like functions grouped close to the θ_T value typical for $m = 1$, the existence of belief relaxation leads to distributions of width comparable to the original values of σ_{0j} centred

Figure 3: **No filtering applied, perfect memory.** Time evolution of the averages of beliefs of the three groups of agents. Rightists: green, centrists: black, leftists: blue. Thin line show evolution of the average belief $\langle \theta \rangle_j(t)$ for individual agents j . Thick lines show the evolution of the ensemble averages for each group of the agents $\langle \theta \rangle_G(t)$. Without filtering, the truth-focused information eventually leads all agents to adopt the θ_T centred beliefs. The process for our choice of $\theta_T = 0.6$ is, of course, the easiest for rightists, who start with beliefs close to this value. However, eventually all groups achieve consensus. The case of leftists (initially holding views opposing θ_T) is quite revealing: agents with initially large tolerances (large initial σ_{0j}) accept truthful information quickly; agents with very focused initial opinions (small values of σ_{0j}) hold on for longer and then very quickly join the majority. Simulations that use the $p = 0.3$ value show the periods in which the individual opinions remain unchanged (indicated by the flat segments of the thin lines).

exactly at θ_T . Thus, while the final ensemble average may be similar, the underlying structure of the individual beliefs is quite different.

Case 2: Individual confirmation bias filter of true information, perfect memory ($m = 1$).

- 4.5** We shall start the analyses of the filtering with the **confirmation bias filter**. There are two reasons for this choice. The first is that confirmation bias is widely recognized in psychological literature, so it ‘deserves’ thorough treatment in the ABM framework. The second reason is the relative simplicity of the filter effects. Suppose that the information flow on which the filter acts is non-specific (i.e., uniform). If the initial belief distribution is given by a Gaussian function with the standard deviation σ , then the application of the same function acting as the likelihood filter would lead to the posterior belief in the Gaussian form, but with σ decreased by a factor of $\sqrt{2}$. A repeated information processing would eventually lead to a Dirac delta-like belief distribution. In other words, **repeated application of confirmation bias narrows and freezes one’s own opinions**.
- 4.6** The simulation set-up for the case of confirmation bias filtering of true information is relatively simple. At every time step, with probability p each agent uses its current belief $X_j(t)$ as the pure filter. In this case the final likelihood function is defined by

$$FL(\theta) = (fX_j(\theta, t) + (1 - f)U)S_T(\theta), \quad (6)$$

where we use the filter effectiveness f as parameter. Small f means that the filter is ‘diluted’. As before, with probability $1 - p$, the agent does not process the information. In this section we focus on situations $m = 1$, when the agent not processing new information simply retains its previous belief distribution. We use here a fixed value of $p = 0.3$.

Figure 4: **No filtering applied, perfect memory.** Evolution of the average mean value of the belief $\langle \theta \rangle_L(t)$ over the group of ‘leftists’ due to ‘truth-related’ information stream. Time is measured in interactions per agent. Black-red curves represent various values of the p parameter value. Decreasing the probability of information-carrying encounters (smaller p) makes the evolution of beliefs slower. The pure Bayesian evolution ($p = 1$) very quickly (in less than a 1000 time steps) leads to the distribution of beliefs centred around the preferred ‘true’ value θ_T . For $p < 1$ the evolution of $\langle \theta \rangle_L(t)$ is stretched in time proportionally to p . Rescaling time to $t' = pt$ shows the invariant shape of the evolution (inset).

Figure 5: **No filtering applied, overview of memory effects.** Time evolution of the leftist group ensemble average $\langle \theta \rangle_L(t)$ for $p = 0.4$ and various values of the memory factor m . Even a relatively small memory loss ($m = 0.99$), i.e., low, but non-zero levels of broadening, speeds up the transition of $\langle \theta \rangle_L$ to the true value θ_T . In other words, agents with belief distributions which are (however seldom) reset to more ‘broadminded’ values are more likely to learn from the ‘truth-related’ information source.

4.7 The relative importance of the confirmation bias filter (progressively narrowing the belief distributions) and the information source depends on the value of the filtering effectiveness factor f . If the filter is used without ‘dilution’ ($f = 1$), the individual beliefs coalesce to a delta-like form in less than 10-time steps and the average beliefs of each group remains practically frozen (left panel in Figure 6). Thus, despite the availability of true information, the centrist and leftist groups remain unconvinced by the information source and keep their beliefs. For much smaller, but still non-negligible value of $f = 0.2$, we can observe some change in group averages (more pronounced for the leftist group, where the dissonance between the initial views and true information is the largest, middle panel in Figure 6).

4.8 It is only for very small values of f that the final distributions of beliefs begin to converge towards the truth-related consensus. Even for $f = 0.05$ there is a sizeable gap between the rightists and the centrists and the

leftists.

Figure 6: **Confirmation bias filtering, perfect memory.** Time evolution of average beliefs $\langle \theta \rangle_j(t)$ (thin lines) and the group averages $\langle \theta \rangle_G(t)$ (thick lines) for the three groups of agents using confirmation bias, for three values of the filter effectiveness $f = 1, 0.2$ and 0.01 . The value of the information processing probability is $p = 0.3$. Decreasing the effectiveness of the confirmation bias filter delays the time at which the individual opinion distributions become fixed and delta-like, shown in the figure as thick horizontal lines. In some cases, we observe jumps in the opinion, typical for discrete Bayesian updates.

Figure 7: **Confirmation bias filtering, perfect memory.** Dependence of the final value of $\langle \theta \rangle_G$ for the three groups as functions of filtering effectiveness f for the confirmation bias filter. Note that the convergence of opinions near the true value requires very weak filtering ($f < 0.02$).

4.9 Figure 7 presents the dependence of the ensemble averaged values of the average belief for each of the three groups $\langle \theta \rangle_G$ on the filtering effectiveness f . For f close to 1, the truth-related information is almost totally filtered out by confirmation bias, agents quickly evolve to fixed, delta-like belief distributions. For medium values ($0.3 < f < 0.9$) the rightists and the centrists show no effects, but the leftists' final average opinions shift towards θ_T . For small values of the filtering effectiveness ($f < 0.1$) the opinions of the three groups begin to converge, but getting close to the consensus requires very small values of f (of the order of 0.02 or less).

Case 2a: Individual confirmation bias filter of true information, broadening of beliefs due to imperfect memory ($m < 1$).

- 4.10** The confirmation bias filter very quickly leads to extreme narrowing of the individual belief distributions (for the fully effective case $f = 1$ this happens after a few tens of interactions). This suggests that the inclusion of the broadening mechanisms due to memory loss in the case of the confirmation bias filter, might have more significant effects than in the case of unfiltered information processing.
- 4.11** In the case of the $S_T(\theta)$ information source, the effects of memory imperfection (opinion broadening) are most clearly seen by the behaviour of the leftist group, because this group is the furthest from the ‘true’ value θ_T . The change is best visualized when we look at the time evolution of the group ensemble average beliefs $\langle \theta \rangle_L$ (Figure 8). The presence of the indeterminacy reset due to imperfect memory causes individual opinion distributions to retain some component of broader beliefs and facilitates their shift due to the influence of the information source. The process $\langle \theta \rangle_L(t) \rightarrow \theta_T$ is quite fast, on the order of a few hundred time steps when $m \leq 0.3$, but slows down for higher values of m . Above $m \approx 0.9$ (i.e., for almost perfect memory) the narrowing of the individual opinion distributions dominates and the group average remain close to their initial values.

Figure 8: **Confirmation bias filtering, memory effects.** Time dependence of the group average value of opinion distribution for the leftist group $\langle \theta \rangle_L$, for various values of memory parameter m , for f equal to 1. Reducing the value of m allows broadening of the individual opinions and changes their evolution. In consequence, all agents become ‘convinced’ by the information source and accept θ_T as the centre of their belief distributions and the group average for m is smaller than certain value (when there is enough broadening). The conviction process is fastest for the lowest values of m . On the other hand, for $m > 0.9$ the agents’ belief distribution remains frozen, which means that the whole system would exhibit significant polarization, despite many interactions with the information source.

- 4.12** The transition, as function of m , between the polarized state at large enough m and the consensus, for smaller m values, is rather abrupt. Figure 9 presents the dependence of the $\langle \theta \rangle_L(t)$ values on m , for two values of the filter effectiveness $f = 1$ and 0.5 for three time snapshots, $t = 1000$, 10000 and 50000. Increasing the time leads to a step-like transition between conditions preserving the polarization and those leading to the consensus.

Case 3: Politically Motivated Reasoning filter

- 4.13** In contrast with the confirmation bias, **the PMR filter is assumed to depend on the current beliefs of the in-group, treated as a whole.** In its simplest version, we assume that any agent knows perfectly the ensemble averaged belief distribution of its in-group $X_G(\theta, t)$, and uses it as a filter for information processing. The filter is dynamical, because as the individual agents change their beliefs, so does the average for the group. One can imagine more advanced and realistic versions of the PMR filter. For example, instead of averaging the in-group opinions (or their subset ‘visible’ to the agent), it could base its motivation on documents describing ‘how a member of the group should think/behave’, such as the party manifestos mentioned before. Or use external reports describing the current ‘official’ position of the group on certain issues.
- 4.14** As in the previous sections we focus on the truth-related information source $S_T(\theta)$ and assume that $p = 0.3$. Our focus is, therefore, the role of the filter effectiveness f in the evolution of the group belief distributions. The current section considers the case of agents with perfect memory ($m = 1$).

Figure 9: **Confirmation bias filtering, memory effects.** Dependence of the value of leftist group ensemble average $\langle \theta \rangle_L(t)$ as function of the memory parameter m , for two values of the filtering effectiveness $f = 1$ and $f = 0.5$, and for three time values, $t = 1000$, 10000 and 50000 steps. For small m values, the group average converges on the true value $\theta_T = 0.6$. For large m values (better memory, i.e. lesser role of the broadening) the beliefs remain largely unchanged. Increasing the time t , at which we measure the $\langle \theta \rangle_L(t)$, makes the transition between two regimes (preserving the original opinions or accepting the true value) less gradual as function of the memory fidelity parameter m .

- 4.15** We shall start with Figure 10, which corresponds directly to the results for the confirmation bias filter (Figure 7). For very small values of f the averaged beliefs converge on the true value, as the information source ‘gets through’, thanks to the uniform part of the filter. On the other hand, for $f \approx 1$, the PMR filtering mechanism effectively freezes the group opinions. For the two groups which are initially closer to the true opinion θ_T , namely the rightists $\langle \theta \rangle_R$ and centrists $\langle \theta \rangle_C$, the fixed value remains unchanged as we lower f , and for very small values of f it changes gradually, resembling the behaviour for the confirmation bias filter. For the leftists, however, instead of a continuous change observed in the confirmation bias case we observe a discontinuous transition at certain value $f_{crit} = 0.43$ (for the current set agents and $p = 0.3$).
- 4.16** To understand this discontinuity we have to look into the details of the evolution of the individual belief distributions. The individual belief distributions $X_j(\theta, t)$, collected for f just above the transition value ($f = 0.43$) and below it ($f = 0.42$) evolve in very different ways. The initial evolution ($t < 10$) is driven by the interplay of the asymmetry of the information source (favouring positive values of θ) and the PMR filter. It leads to formation of two attractors, around which the individual agents group: one close to the upper end of the original leftist domain (around $\theta = -0.5$) and the second, corresponding to partially ‘convinced’ agents, located around $\theta = 0.1$. The decrease of filter effectiveness f increases the number of agents in the latter group. Because the ensemble averaged belief distribution enters the process for the next iteration, for $f < 0.42$, a positive feedback mechanism leads to the eventual dominance of the convinced group. On the other hand, for $f > 0.43$ the size of the convinced group is too small to persist, and eventually all agents retain or revert to their leftist stance.
- 4.17** The results for the Politically Motivated Reasoning filter were obtained using an assumption that the composition of the group to which an agent looks for the belief guidance remains unchanged. The simulations assume that each agent considers the whole group, defined in the initial input files, to calculate the ensemble averaged belief distribution $X_G(\theta, t)$, which would be used as the filter. This leads to the case when the more flexible agents, who have shifted their opinion can eventually pull the whole group with them (for small enough f values).
- 4.18** Such assumption might be criticized from a sociological point of view. In a situation where the belief systems of the flexible and inflexible agents have very little overlap, one could expect that each of the sub-groups would

Figure 10: **Politically Motivated Reasoning filtering, perfect memory.** Dependence of the final value of $\langle \theta \rangle_G$ for the three groups, as functions of filtering effectiveness f for the PMR filter. For $f \gtrsim 0.43$ the averages are almost independent of f . At $f \approx 0.43$ (marked by the red ellipse), the $\langle \theta \rangle_L$ shows a large jump towards the θ_T value, the result effectively turns leftists into centrists. For very small values of the filtering effectiveness ($f < 0.1$) opinions of all three groups converge on the true value $\theta_t = 0.6$.

restrict their PMR filter to the group of the currently like-minded agents. In other words, the flexibles, who have moved away from the initial group average, would be rejected by the less flexible agents, as traitors of the cause, and disregarded when calculating the PMR filter. The obvious result would be a split of the initial group, occurring within just a few filtered iterations (somewhere between $t = 25$ and $t = 50$). In this approach it would be useful to change the simulation measurements from the group averages of belief $\langle \theta \rangle_G$ to the **numbers** of inflexibles, unconvinced by the information, and the agents who have shifted their beliefs. Such a dynamical group composition model variant shall be the topic of later works.

Case 3a: PMR filter with imperfect memory opinion broadening ($m < 1$)

- 4.19** Figure 11, which presents the results of the PMR filter for $m = 0.5$, confirms that these expectations are indeed true. Instead of the discrete jump seen for the leftist group in the unmodified $m = 1$ case (Figure 10), we can observe smooth changes of all group final averages of beliefs $\langle \theta \rangle_G$ (Figure 11). Moreover, a full consensus is reached for finite (although small) values of f . An additional difference in the simulations for the imperfect memory PMR filter from all cases considered so far, is that simulation runs converge to somewhat different configurations. We have indicated this as error bars in Figure 11.
- 4.20** The roughly linear dependence of $\langle \theta \rangle_L$ on f , for $f > 0.1$, results from the increased individual opinion flexibility introduced by the mixture of broad-minded components of individual beliefs treated as priors. To better understand this, we have studied the dependence of $\langle \theta \rangle_L$ on the memory factor m for fixed values of f . The results are shown in Figure 12. In the case of relatively effective PMR filter ($f = 0.7$ and $f = 1.0$) there are two distinct regimes of system behaviour. Above a certain threshold value $m_T(f)$, there is only a weak, linear dependence of $\langle \theta \rangle_L$ on m , mostly due to individual belief shiftings during a few initial time steps, which quickly become frozen. On the other hand, for m smaller than $m_T(f)$, all agents shift their opinions in accordance with the information source, moving eventually to centrist and rightist positions. The value of $m_T(f)$ is only approximate, as a consequence of differences between individual simulation runs, due to the finite size of the system.

Figure 11: **Politically Motivated Reasoning filtering, $m = 0.5$ (broadening of belief distribution).** Dependence of the final value of $\langle \theta \rangle_G$ for the three groups, as functions of filtering effectiveness f for the PMR filter with imperfect memory $m = 0.5$. The broadening of individual belief distribution due to imperfect memory restores the almost linear dependence of the ensemble average value of opinion for the leftist group. The resulting opinion distribution for the leftist group $\langle \theta \rangle_L$, for $f > 0.6$, shows sizeable differences between individual simulation runs, which are indicated by error bars.

Figure 12: **Politically Motivated Reasoning filtering, memory effects.** Dependence of the final value of $\langle \theta \rangle_G$ for the leftist group on the memory factor m , for two values of the filtering effectiveness f for the PMR filter. Dots show results of the individual simulation runs. Red ellipses indicate the regions close to the threshold value of m , at which the behaviour of the system changes. Decreasing the memory quality from the perfect case ($m = 1$) leads initially to very slight, linear shift in the $\langle \theta \rangle_L$ value, attributable to belief changes in the first few interactions. Below the threshold value (which depends on f) the group opinion average grows to approach the true value of θ_T for $m < 0.1$. The black lines are separate best fits of linear function (for m greater than a threshold value) and quadratic function for m smaller than it.

Discussion

- 5.1** The simulations presented here are based on drastically simplified assumptions, i.e., a single source of information, with a fixed $S_T(\theta)$ distribution and consistently repeated filter. These simplifications clearly suggest the directions of further work: dealing with conflicting information sources, combinations of different types of filters, transient phenomena to describe immediate reactions to the exposure of news. Another planned model extension is related to modelling the possible dynamics nature of group norm based PMR filters, with the filter shape depending on an average perceived opinion distribution of any changing group. When opinions within a group which is initially treated as homogeneous begin to diverge, it is quite likely that the very definition of the group would change. Agents could redefine the criteria *who they count as the members of the in-group*, treating those with sufficiently different belief distributions as outsiders (possibly with a negative emotional label of traitors). Such a move would dynamically redefine the perceived in-group standards and norms. The resulting change in the PMR filter could change the model dynamics from opinion shifts to changes in group sizes and identification.
- 5.2** The model proposed in this work belongs to the ‘**rich feature agent**’ or ‘**complex agent**’ categories, in contrast to the simplified ‘spinon’ models. To examine the possibilities of this approach, we have focused on a system in which agents repeatedly react to an unchanging, single external information source. This has allowed us to discover certain regularities and to understand the roles of model parameters.
- 5.3** The same general framework of biased processing of information may be used in more complex environments. It can cover agents interacting among themselves in arbitrarily chosen social networks. In such a scenario, the input information would be generated by one of agents (a sender), and would be received and evaluated using filtering mechanisms and biases by other agents (recipient(s)). Each recipient would then update its opinion (as described by the belief distribution), and if applicable for the bias type, also the filter function. Of course, it is possible to reverse the roles of agents and to allow bi/directional communication. Because the filters used by the communicating agents may be different, the interaction process may be asymmetric. It is also possible to combine agent-to-agent interactions with the influences of external information sources, and to create a truly complex model approximating a real society or its fragment. Moreover, the interacting agents may differ in their characteristics. For example, the issue may be considered as more or less important (making the filtering efficiency specific to each agent and time $f \rightarrow f_j(t)$). Similarly, the resources available to each agent may be different (access to information sources and frequency of events, for example), which would make the modelling even more complicated.
- 5.4** In addition to the complexity of the individual agent’s characteristics, the model could be expanded to include observations that in most cases we should not consider individual issues, but rather belief systems, combining larger number of beliefs, all connected and evaluated together.
- 5.5** Lastly, especially in the case of short term, transient changes, the possibilities of manipulation of filters by outside agencies, offer a very interesting and important future research direction. Such investigations should cover both manipulations that increase polarization (partisan information sources and the reliance on emotional context of the information) as well as the efforts in the opposite direction – to detect and combat manipulative influences. The latter are especially important to enhance meaningful dialogue in our already highly polarized society.
- 5.6** Our results should not be considered as anything more than a ‘proof of concept’ of the proposed framework. To describe real world situations, one should expand it to take into account the aspects discussed above. In doing so for a system that is too complex (e.g., political situations in any given country), we risk getting lost in the assignment of the relative roles and strengths of various mechanisms. For this reason, the suggested way forward is to couple modelling with small scale psychological laboratory experiments, for which various factors could be controlled and monitored. Such work facilitates the ‘calibration’ of a model’s parameters and could improve definitions of the filtering functions for various types of information processing. It would also bridge ABM and psychology approaches in a framework understandable to both communities.

Other types of information filters

- 5.7** The way in which information received from various sources is evaluated and used to form new beliefs, depends not only on sources, but also on the goals of a person. These goals may allow us to construct rules that would create and update the information filters. In some cases, they would be independent of a person’s characteristics, in other cases they would depend on them, which would make the process of belief modification self-referential. Below is a partial list of the filter types that could be used in the future complex agent-based mod-

elling. The filters are distinguished by their origin (internal to the person or external), dependence on some objectively measurable characteristics, possibility of an orchestrated manipulation and, finally, normative value.

- **Memory priming/availability filter.** This is an example of an internal filter (similarly to confirmation bias), which however, is much more easily manipulated. This is because the confirmation bias compares new information with currently held beliefs, which may be quite deeply ingrained, especially if they depend on moral foundations (Jost et al. (2003); Haidt (2007, 2012a,b)). In contrast, the availability filter acts via additional attention given to facts that are quickly accessible. Thanks to various forms of priming, its effects may be effectively stimulated and steered by outside influence: our peers or the media (Tversky & Kahneman (1974, 1993); Sunstein (2006)). In terms of an ABM approach, such filters could be approximated, for example, by the shape of the previously encountered information source.
- **Simplicity/attention limit filter.** This is another internal filter, related to the culturally and technologically driven change in the way external information is processed. Due to the information deluge, there is an increasing dominance of short forms of communication, especially in the Internet based media: WEB pages, Internet discussions, social media (Djamasbi et al. (2011, 2016)). The simplification (or over-simplification) of important issues, necessary to fit them into short communication modes, may act against beliefs that are not disposed to such simplification. This part of the filter acts at the creation side of the information flow. Decreasing attention span and capacity to process longer, argumentative texts act as another form of filter, this time at the reception end of the flow. There are numerous forms of psychological bias related to and leading to such filtering, from venerable and accepted heuristics (like Occam's razor), through the law of triviality and bike-shed syndrome (Parkinson (1958)), to a total disregard for too complex viewpoints (Qiu et al. (2017)). Together, these tendencies can create a filter favouring information that is easily expressed in a short, catchy, easy to memorize form. There is no simple universal form of the filter for the ABM approach, because in different contexts different beliefs might be easier to be expressed in the most simple way.
- **Emotional filter.** Some topics, contexts and communication forms may depend, in their processing, on their affective or emotional content. This can create a processing filter, for example one that favours extreme views, as they are typically more emotional than the consensus oriented, middle-of-the-road ones. Emotionally loaded information elicits stronger response and longer lasting effects (Hatfield et al. (1993); Haidt (2001); Barsade (2002); Allen et al. (2005); Clore & Huntsinger (2007); Berger & Milkman (2010); Nielek et al. (2010); Sobkowicz & Sobkowicz (2010); Chmiel et al. (2011b,a); Reifen Tagar et al. (2011); Thagard & Findlay (2011); Sobkowicz (2012); Bosse et al. (2013)). The specific form of the filter depends on the mapping of the belief range and the associated emotional values. Furthermore, the emotional filter may depend on the current agent belief function, e.g., anger directed at information contrary to the currently held beliefs, or at a person who acts as the source of the information.
- **Algorithmic filters.** An increasing part of information reaching us comes from the Internet services such as our own social media accounts, personalized search profiles etc. Service providers organize and filter the content that reaches us often without our knowledge that any filter exists; and even more often without the knowledge of how it works. These external algorithmic filters, shaping our perception, not only skew opinions but more importantly, they often limit the range of topics we are aware of and the opinions related to them (Pariser (2011); Albanie et al. (2017)). In some cases, the effect of an algorithmic filter is similar to its internal confirmation bias (e.g., the search engine prioritizes the results based on already recognized preferences of the user). In other cases, the machine filter may deliberately steer the user away from certain information, based on decisions unrelated to the particular user, fulfilling the goals of some other third party.

Time dependency considerations

- 5.8 The choice of the right simulation-to-reality time scaling may depend on the way we define the information processing events. On one hand, we could consider only major news and real-world occurrences, such as crucial election stories and events. In such cases, the number of opinion shaping encounters could be treated as relatively small, certainly not in the range of tens of thousands or thousands per month. In such a view, time periods between information processing events are long enough to allow uncertainty to reset.
- 5.9 At the other end of the spectrum is the vision, in which our beliefs are shaped by a continuous stream of events, differing in their source type, intensity, repetition and many other characteristics. Some of these would originate from external sources, characterized by relatively stable views and opinions (biased or unbiased at the

source), while other events could originate from more or less random encounters with other people or observations of ostensibly small importance. In such microscopic approach, the number of the events could be very large.

- 5.10 The focus of this work was on the long-term effects of a single type of an information source, interspersed with periods when an individual belief structure may become less certain. The goal was to construct a general framework for filtered information processing and see if such an approach can yield 'reasonable results', by which we mean, depending on the situation, conditions leading to a general consensus, or for other conditions, persistent disagreement and polarization. The results have shown that the model can indeed, produce these results under simple manipulation of a few key parameters.
- 5.11 The question of the 'right' time-scale for opinion change can not be resolved by such a qualitative, simplified model. Among the unknowns there is the effectiveness of the update process and filtering, the memory imperfection related uncertainty reset scale, and elements omitted in the current model, for example differences in the intensity of particular events. A more realistic model should be based on psychological studies – which would, hopefully, also provide suggestions as to whether we should focus on the effects of a few (few tens? hundreds?) information processing events or to look at the stable or quasi-stable states reached after thousands of microscopic events.

Possibility of manipulation of the Politically Motivated Reasoning Filter

- 5.12 Current political developments in many democratic societies show dramatically increasing levels of polarization, covering the general public and the media (Baldassarri & Bearman (2007); Fiorina & Abrams (2008); Bernhard et al. (2008); Stroud (2010); Prior (2013); PEW (2014); Tewksbury & Riles (2015)). In many countries the chances of reaching a state in which a rational discussions between conflicted groups (not to mention working out a sensible compromise) seems almost impossible. Recent US presidential elections provide an obvious example, but the seemingly irrevocable split exists in many other aspects, sometimes with division lines not parallel to political ones. A good example of such a split is the existence and (in many countries) growth of the anti-vaccination movements (Streefland (2001); Davies et al. (2002); Wolfe & Sharp (2002); Leask et al. (2006); Blume (2006); Nelson (2010); Kata (2010); Betsch (2011); Betsch & Sachse (2013); Ołpiński (2012); McKeever et al. (2016); Hough-Telford et al. (2016)), which are not strictly 'politically' aligned. Efforts to convince vaccination opponents are quite unsuccessful, regardless of the approach used. Similar problems occur in more politicized issues. This applies to the cases where suitable evidence is available, for example in controversies over gun control policies, climate change, GMO, nuclear energy, and in cases where beliefs and opinions are largely subjective, such as evaluations of specific politicians (e.g., Hillary Clinton or Donald Trump).
- 5.13 The US presidential election in 2016, with its increasing role of social media as information sources, has focused attention on yet another form of 'at source' information manipulation: i.e., **fake news**. The relative ease of creating false information, in some cases supported by manipulated images, voice and video recordings, to post it on-line and to create a web of self-supporting links allows the perpetrator to spread such news. The trust associated with social networks (for example Facebook or twitter links) makes spreading such information faster – especially if the fake news is designed to pass through the most common information filters.
- 5.14 The difficulty in minimizing polarization may be partially attributed to the cognitive biases and motivated information processing described in this paper. Filtering-out information may be very effective in keeping a person's beliefs unchanged. In fact, certain cognitive heuristics are evolved to provide this stability (e.g. the confirmation bias). This makes the task of bridging gaps between polarized sections of our societies seem impossible. Still, as Kahan has noted, some filtering mechanisms may be more flexible than others.
- 5.15 A good example is provided by comparison of the confirmation bias and PMR. Kahan (2016a) notes that in some cases PMR may be confused with the confirmation bias: *"Someone who engages in politically motivated reasoning will predictably form beliefs consistent with the position that fits her predispositions. Because she will also selectively credit new information based on its congeniality to that same position, it will look like she is deriving the likelihood ratio from her priors. However, the correlation is spurious: a 'third variable'—her motivation to form beliefs congenial to her identity – is the 'cause' of both her priors and her likelihood ratio assessment."* Kahan notes the importance of the difference: if the source of the filter is 'internal' (confirmation bias), we have little hope in modifying it. On the other hand, if the motivation for filtering is related to perceptions of in-group norms, **opinions may be changed if the perception of these in-group norms changes**. Re-framing issues in a language that conforms to specific in-group identifying characteristics or providing information that certain beliefs are 'in agreement' with the value system of the in-group and/or majority of its members, would change the PMR

filtering mechanism. Through this change, more information could be allowed through, changing the Bayesian likelihood function, and, eventually, changing the posterior beliefs.

References

- Adamic, L. & Glance, N. (2005). The political blogosphere and the 2004 US election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery*, (pp. 36–43). Chicago (IL): ACM.
- Albanie, S., Shakespeare, H. & Gunter, T. (2017). Unknowable manipulators: Social network curator algorithms. *arXiv preprint arXiv:1701.04895*
- Allen, C., Machleit, K., Kleine, S. & Notani, A. (2005). A place for emotion in attitude models. *Journal of Business Research*, 58(4), 494–499
- Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, 193(5), 31–35
- Asch, S. E. & Guetzkow, H. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. S. Guetzkow (Ed.), *Groups, Leadership, and Men: Research in Human Relations*, (pp. 222–236). Oxford: Carnegie Press.
- Baldassarri, D. & Bearman, P. (2007). Dynamics of political polarization. *American Sociological Review*, 72(5), 784
- Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, 47(4), 644–675
- Ben-Naim, E., Frachebourg, L. & Krapivsky, P. L. (1996). Coarsening and persistence in the voter model. *Physical Review E*, 53(4), 3078–3087
- Benson, B. (2016). Cognitive bias cheat sheet
URL <https://betterhumans.coach.me/cognitive-bias-cheat-sheet-55a472476b18>
- Berger, J. & Milkman, K. L. (2010). Social transmission, emotion, and the virality of online content. Tech. rep., Wharton School, University of Pennsylvania
- Bernardes, A. T., Costa, U. M. S., Araujo, A. D. & Stauffer, D. (2001). Damage spreading, coarsening dynamics and distribution of political votes in Sznajd model on square lattice. *International Journal of Modern Physics C*, 12(2), 159–168
- Bernhardt, D., Krasa, S. & Polborn, M. (2008). Political polarization and the electoral effects of media bias. *Journal of Public Economics*, 92(5-6), 1092–1104
- Betsch, C. (2011). Innovations in communication: The Internet and the psychology of vaccination decisions. *Euro Surveill*, 16, 17
- Betsch, C. & Sachse, K. (2013). Debunking vaccination myths: Strong risk negations can increase perceived vaccination risks. *Health Psychology*, 32(2), 146
- Blume, S. (2006). Anti-vaccination movements and their interpretations. *Social Science & Medicine*, 62(3), 628–642
- Bond, R. & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952, 1956) line judgment task. *Psychological Bulletin*, 119(1), 111
- Bosse, T., Hoogendoorn, M., Klein, M. C., Treur, J., Van Der Wal, C. N. & Van Wissen, A. (2013). Modelling collective decision making in groups and crowds: Integrating social contagion and interacting emotions, beliefs and intentions. *Autonomous Agents and Multi-Agent Systems*, 27(1), 52–84
- Bullock, J. (2009). Partisan bias and the Bayesian ideal in the study of public opinion. *The Journal of Politics*, 71(03), 1109–1124
- Campante, F. R. & Hojman, D. A. (2010). Media and polarization. Tech. rep., Harvard University, John F. Kennedy School of Government
URL <http://nrs.harvard.edu/urn-3:HUL.InstRepos:4454154>

- Caruso, F. & Castorina, P. (2005). Opinion dynamics and decision of vote in bipolar political systems. *International Journal of Modern Physics C*, 16(09), 1473–1487
- Castellano, C. (2012). Social influence and the dynamics of opinions: The approach of statistical physics. *Managerial and Decision Economics*, 33(5-6), 311–321.
- Castellano, C., Fortunato, S. & Loreto, V. (2009). Statistical physics of social dynamics. *Review of Modern Physics*, 81, 591–646
- Castellano, C., Vilone, D. & Vespignani, A. (2003). Incomplete ordering of the voter model on small-world networks. *EPL (Europhysics Letters)*, 63, 153
- Chmiel, A., Sienkiewicz, J., Thelwall, M., Paltoglou, G., Buckley, K., Kappas, A. & Holyst, J. (2011a). Collective emotions online and their influence on community life. *PLoS ONE*, 6(7), e22207
- Chmiel, A., Sobkowicz, P., Sienkiewicz, J., Paltoglou, G., Buckley, K., Thelwall, M. & Holyst, J. (2011b). Negative emotions boost users activity at BBC forum. *Physica A*, 390(16), 2936–2944.
- Clore, G. L. & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences*, 11(9), 393
- Cox, J. & Griffeath, D. (1986). Diffusive clustering in the two dimensional voter model. *The Annals of Probability*, 14(2), 347–370
- Davies, P., Chapman, S. & Leask, J. (2002). Antivaccination activists on the world wide web. *Archives of Disease in Childhood*, 87(1), 22–25
- Deffuant, G., Amblard, F., Weisbuch, G. & Faure, T. (2002). How can extremism prevail? A study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation*, 5(4), 1
- Deffuant, G., Neau, D., Amblard, F. & Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3, 87–98
- Djamasbi, S., Rochford, J., DaBoll-Lavoie, A., Greff, T., Lally, J. & McAvoy, K. (2016). Text simplification and user experience. In *International Conference on Augmented Cognition*, (pp. 285–295). Berlin/Heidelberg: Springer
- Djamasbi, S., Siegel, M., Skorinko, J. & Tullis, T. (2011). Online viewing and aesthetic preferences of generation Y and the baby boom generation: Testing user web site experience through eye tracking. *International Journal of Electronic Commerce*, 15(4), 121–158
- Fiorina, M. P. & Abrams, S. J. (2008). Political polarization in the American public. *Annual Review of Political Science*, 11, 563–588
- Fortunato, S. & Castellano, C. (2007). Scaling and universality in proportional elections. *Physical Review Letters*, 99(13), 138701
- Galam, S. (2012). *Sociophysics: A Physicist's Modeling of Psycho-Political Phenomena*. Berlin/Heidelberg: Springer
- Galam, S. (2017). The Trump phenomenon, an explanation from sociophysics. *International Journal of Modern Physics B* 31, 1742015
- Galam, S., Chopard, B. & Droz, M. (2002). Killer geometries in competing species dynamics. *Physica A: Statistical Mechanics and its Applications*, 314(1), 256–263
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316(5827), 998–1002
- Haidt, J. (2012a). Left and right, right and wrong. *Science*, 337, 525–526
- Haidt, J. (2012b). *The Righteous Mind: Why Good People Are Divided by Politics and religion*. New York (NY): Vintage

- Hatfield, E., Cacioppo, J. T. & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Science*, 2(3), 96–99
- Hegselmann, R. & Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2
- Holyst, J., Kacperski, K. & Schweitzer, F. (2001). Social impact models of opinion dynamics. *Annual Review of Computational Physics*, IX, 253–273.
- Hough-Telford, C., Kimberlin, D. W., Aban, I., Hitchcock, W. P., Almquist, J., Kratz, R. & O'Connor, K. G. (2016). Vaccine delays, refusals, and patient dismissals: A survey of pediatricians. *Pediatrics*
- Jerit, J. & Barabas, J. (2012). Partisan perceptual bias and the information environment. *Journal of Politics*, 74(3), 672–684
- Jost, J., Glaser, J., Kruglanski, A. & Sulloway, F. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, 129(3), 339–375
- Jost, J. T., Hennes, E. P. & Lavine, H. (2013). Hot political cognition: Its self-, group-, and system-serving purposes. In D. E. Carlston (Ed.), *Oxford Handbook of Social Cognition*, (pp. 851–875). Oxford: Oxford University Press
- Kacperski, K. & Holyst, J. (1999). Opinion formation model with strong leader and external impact: A mean field approach. *Physica A*, 269, 511–526
- Kacperski, K. & Holyst, J. (2000). Phase transitions as a persistent feature of groups with leaders in models of opinion formation. *Physica A*, 287, 631–643
- Kahan, D. M. (2016a). The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it. In R. Scott & S. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences*. Chichester: Wiley
- Kahan, D. M. (2016b). The politically motivated reasoning paradigm, part 2: Unanswered questions. In R. Scott & S. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences*. Chichester: Wiley
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Basingstoke: MacMillan
- Kata, A. (2010). A postmodern Pandora's box: Anti-vaccination misinformation on the internet. *Vaccine*, 28(7), 1709–1716
- Lawrence, E., Sides, J. & Farrell, H. (2010). Self-segregation or deliberation? Blog readership, participation, and polarization in American politics. *Perspectives on Politics*, 8(01), 141–157
- Leask, J., Chapman, S., Hawe, P. & Burgess, M. (2006). What maintains parental support for vaccination when challenged by anti-vaccination messages? A qualitative study. *Vaccine*, 24(49), 7238–7245
- Lindenberg, S. (2001). Social rationality as a unified model of man (including bounded rationality). *Journal of Management and Governance*, 5(3), 239–251
- Lindenberg, S. (2010). Why framing should be all about the impact of goals. In P. Hill, F. Kalter, J. Kopp, C. Kroeber & R. Schnell (Eds.), *Hartmut Essers Erklärende Soziologie*, (pp. 53–79). Frankfurt am Main: Campus
- Martins, A. C. (2008). Continuous opinions and discrete actions in opinion dynamics problems. *International Journal of Modern Physics C*, 19(04), 617–624
- Martins, A. C. (2009). Bayesian updating rules in continuous opinion dynamics models. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(02), P02017
- Martins, A. C. (2014). Discrete opinion models as a limit case of the coda model. *Physica A: Statistical Mechanics and its Applications*, 395, 352–357
- Martins, A. C. R. & Kuba, C. D. (2010). The importance of disagreeing: Contrarians and extremism in the coda model. *Advances in Complex Systems (ACS)*, 13(05), 621–634
- McKeever, B. W., McKeever, R., Holton, A. E. & Li, J.-Y. (2016). Silent majority: Childhood vaccinations and antecedents to communicative action. *Mass Communication and Society*, 19(4), 476–498

- Nelson, K. (2010). Markers of trust: How pro-and anti-vaccination web sites make their case. Tech. Rep. 1579525, SSRN
- Ngampruetikorn, V. & Stephens, G. J. (2016). Bias, belief, and consensus: Collective opinion formation on fluctuating networks. *Physical Review E*, 94(5), 052312
- Nielek, R., Wawer, A. & Wierzbicki, A. (2010). Spiral of hatred: Social effects in Internet auctions. Between informativity and emotion. *Electronic Commerce Research*, 10, 313–330.
- Nowak, A. & Lewenstein, M. (1996). Modeling social change with cellular automata. In R. Hegselmann, U. Mueller & K. G. Troitzsch (Eds.), *Modelling and Simulation in the Social Sciences from a Philosophy of Science Point of View*, (pp. 249–285). Dordrecht: Kluwer
- Nowak, A., Szamrej, J. & Latané, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review*, 97(3), 362–376
- Nyczka, P. & Sznajd-Weron, K. (2013). Anticonformity or independence? – Insights from statistical physics. *Journal of Statistical Physics*, 151, 174–202
- Ołpiński, M. (2012). Anti-vaccination movement and parental refusals of immunization of children in USA. *Pediatrica Polska*, 87(4), 381–385
- Opaluch, J. J. & Segerson, K. (1989). Rational roots of ‘irrational’ behavior: New theories of economic decision-making. *Northeastern Journal of Agricultural and Resource Economics*, 18(2), 81–95
- Palombi, F. & Toti, S. (2015). Voting behavior in proportional elections from agent-based models. *Physics Procedia*, 62, 42–47
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. London: Penguin
- Parkinson, C. N. (1958). *Parkinson's Law: The Pursuit of Progress*. London: John Murray
- Pew Research Center (2014). Political polarization in the American public.
URL <http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/>
- Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, 16, 101–127
- Qiu, X., Oliveira, D. F., Shirazi, A. S., Flammini, A. & Menczer, F. (2017). Limited individual attention and online virality of low-quality information. *arXiv preprint arXiv:1701.02694*
- Reifen Tagar, M., Federico, C. & Halperin, E. (2011). The positive effect of negative emotions in protracted conflict: The case of anger. *Journal of Experimental Social Psychology*, 47(1), 157–164
- Sabatelli, L. & Richmond, P. (2003). Phase transitions, memory and frustration in a Sznajd-like model with synchronous updating. *International Journal of Modern Physics C*, 14, 1223–1229
- Sabatelli, L. & Richmond, P. (2004). Non-monotonic spontaneous magnetization in a Sznajd-like consensus model. *Physica A: Statistical Mechanics and Its Applications*, 334(1), 274–280
- Shiller, R. (1995). Conversation, information, and herd behavior. *The American Economic Review*, 85(2), 181–185
- Slanina, F. & Lavicka, H. (2003). Analytical results for the Sznajd model of opinion formation. *European Physical Journal B - Condensed Matter*, 35(2), 279 – 288
- Sobkowicz, P. (2009). Modelling opinion formation with physics tools: Call for closer link with reality. *Journal of Artificial Societies and Social Simulation*, 12(1), 11
- Sobkowicz, P. (2010). Effect of leader's strategy on opinion formation in networked societies with local interactions. *International Journal of Modern Physics C (IJMPC)*, 21(6), 839–852.
- Sobkowicz, P. (2012). Discrete model of opinion changes using knowledge and emotions as control variables. *PLoS ONE*, 7(9), e44489.
- Sobkowicz, P. (2013a). Minority persistence in agent based model using information and emotional arousal as control variables. *The European Physical Journal B*, 86(7), 1–11.

- Sobkowicz, P. (2013b). Quantitative agent based model of user behavior in an internet discussion forum. *PLOS One*, 8(12), e80524.
- Sobkowicz, P. (2016). Quantitative agent based model of opinion dynamics: Polish elections of 2015. *PloS one*, 11(5), e0155098
- Sobkowicz, P. & Sobkowicz, A. (2010). Dynamics of hate based internet user networks. *The European Physical Journal B*, 73(4), 633–643.
- Stauffer, D. (2001). Monte Carlo simulations of Sznajd models. *Journal of Artificial Societies and Social Simulation*, 5(1), 4
- Stauffer, D. (2002). Sociophysics: The Sznajd model and its applications. *Computer Physics Communications*, 146(1), 93–98
- Stauffer, D. & de Oliveira, P. M. C. (2002). Persistence of opinion in the Sznajd consensus model: Computer simulation. *The European Physical Journal B-Condensed Matter*, 30(4), 587–592.
- Streefland, P. H. (2001). Public doubts about vaccination safety and resistance against vaccination. *Health Policy*, 55(3), 159–172
- Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, 60(3), 556–576
- Suen, W. (2004). The self-perpetuation of biased beliefs. *The Economic Journal*, 114(495), 377–396
- Sugden, R. (2003). Reference-dependent subjective expected utility. *Journal of Economic Theory*, 111(2), 172–191
- Sunstein, C. (2000). Deliberative trouble? Why groups go to extremes. *The Yale Law Journal*, 110(1), 71–119
- Sunstein, C. R. (2002). *Risk and Reason: Safety, Law, and the Environment*. New York, NY: Cambridge University Press
- Sunstein, C. R. (2006). The availability heuristic, intuitive cost-benefit analysis, and climate change. *Climatic Change*, 77(1-2), 195–210
- Sunstein, C. R., Bobadilla-Suarez, S., Lazzaro, S. C. & Sharot, T. (2016). How people update beliefs about climate change: Good news and bad news. Tech. Report SSRN 2821919
- Sznajd-Weron, K. & Sznajd, J. (2000). Opinion evolution in closed community. *International Journal of Modern Physics C*, 11, 1157–1166
- Tafari, S., Gallone, M., Cappelli, M., Martinelli, D., Prato, R. & Germinario, C. (2014). Addressing the anti-vaccination movement and the role of HCWs. *Vaccine*, 32(38), 4860–4865
- Tewksbury, D. & Riles, J. M. (2015). Polarization as a function of citizen predispositions and exposure to news on the internet. *Journal of Broadcasting and Electronic Media*, 59(3), 381–398
- Thagard, P. & Findlay, S. (2011). Changing minds about climate change: Belief revision, coherence, and emotion. In E. Olsson & S. Enqvist (Eds.), *Belief Revision Meets Philosophy of Science: Logic, Epistemology, and the Unity of Science*, (pp. 329–345). Berlin/Heidelberg: Springer
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131
- Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 30
- Tversky, A. & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, 59(4), S251–S278
- Tversky, A. & Kahneman, D. (1993). Probabilistic reasoning. *Readings in Philosophy and Cognitive Science*, (pp. 43–68)
- Tversky, A., Slovic, P. & Kahneman, D. (1990). The causes of preference reversal. *The American Economic Review*, 80(1), 204–217
- Weisbuch, G. (2004). Bounded confidence and social networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38(2), 339–343

- Weisbuch, G., Deffuant, G., Amblard, F. & Nadal, J.-P. (2003). Interacting agents and continuous opinions dynamics. In R. Cowan & N. Jonard (Eds.), *Heterogenous Agents, Interactions and Economic Performance*, vol. 521 of *Lecture Notes in Economics and Mathematical Systems*, (pp. 225–242). Berlin/Heidelberg: Springer
- Wojcieszak, M., Bimber, B., Feldman, L. & Stroud, N. J. (2016). Partisan news and political participation: Exploring mediated relationships. *Political Communication*, 33(2), 241–260
- Wolfe, R. M. & Sharp, L. K. (2002). Anti-vaccinationists past and present. *British Medical Journal*, 325(7361), 430
- Wolfe, R. M., Sharp, L. K. & Lipsky, M. S. (2002). Content and design attributes of antivaccination web sites. *Journal of the American Medical Association*, 287(24), 3245–3248