Whither now, opinion modelers?

1 Introduction

The idea to use the concepts and rigor of physical sciences to describe social phenomena dates back as far as 18th century. One could point to Marquis de Condorcet, who introduced a concept of social arithmetics in 1794. Adolphe Quetelet published *Sur l'homme et le développement de ses facultés, ou Essai de physique sociale* in 1835. Auguste Comte proposed *physique sociale* (social physics) as a new science in 1839, in the fourth tome of his *Cours de philosophie positive* . The arguments were not just purely scientific—applying to social sciences the same requirements as to physics or chemistry was proposed as the means of promoting progress through better understanding. But it was only during the last decades of the 20th century, when advances in physics of complex systems and in computer technologies have allowed actual possibility to apply these ideas.

Often referred to as " sociophysics" [1 , 2], these studies covered a range of social phenomena. Among those, modeling of social processes of opinion formation was among the earliest applications. Using analogies with physics of magnetism, the early models focused on conditions in which society members (compared to atoms in a crystalline lattice) would achieve consensus (equivalent to ferromagnetic transition). Direct mapping of notions from physics, such as temperature, has resulted in model parameters that were incomprehensible for social scientists. One could legitimately ask what is the " social temperature"? How does one measure it? How does limiting the interactions to immediate neighborhood on a periodic lattice correspond to the complexity of human relationships? Moreover, the models usually followed the path of spontaneous

magnetization studies, that is, they started from random initial conditions (corresponding in physical systems to high enough temperatures) and followed the ordering resulting from " cooling down." The obvious problem that social systems are seldom fully randomized was often glossed over. Similarly neglected was the observation that it is by far not obvious that the social systems are in equilibrium state (or even to a steady state).

With very few exceptions, sociophysical models were created without active participation of sociologists or psychologists, and published in journals typically neglected by the social sciences community. A similar gulf separated the modeling community from the practitioners of public opinion measurement.

The later developments have alleviated some of the above problems, bringing models closer to descriptions of real social systems, at least in a qualitative way. The interacting entities corresponding to people and their relationships grew more complex, and a new name, agent-based models (ABMs), grew in popularity. The parallel rise in Internet-based interactions and social networks, which have allowed to observe at least the expressed opinions and to follow the patterns of connections between people and their reactions to various types of messages, has allowed the models to get closer to quantitative description goals, at least for some of the characteristic features. Advances in data-mining techniques allowing to recognize not only topics but also associated sentiments of these digital communications, spurred extensions of the models based on more than a simple one-dimensional opinion range, for example, including emotions. It is interesting to observe that while the focus of the early works was " getting to

consensus" (achieving perfect magnetization), the rise of polarization and persistent conflicts in many societies have prompted some (but not all) researchers to construct models which would result in polarized, disjoint final states.

The rise of populist ideas, in many cases, considered to be dangerous by the political commentators and actively shunned by the traditional political parties (e. g., AfD in Germany or National Front/National Rally in France); growth of polarization and partisanship spanning whole social structures (from social grassroots to the highest institutions, e. g., the U. S. Congress and Senate and even the Supreme Court); the apparent impossibility of combatting various forms of terrorism all have indicated the need for a better understanding of the phenomena driving social opinion changes. The response of the modeling community has been, however, mixed. In the current article, we shall describe not only the history and current status of the agent-based models of opinion dynamics but also the potential directions of development and choices faced by the research community.

2 Three Model Generations

The development of social opinion dynamics models may be roughly divided into three phases. In the first one, very simple models, based on rough, intuitive analogies with certain physical systems were introduced. This phase, which started some thirty years ago, was focused on " discovering" the qualitative similarities between physical phenomena such as spontaneous magnetization and social ones, for example, appearance of a consensus within a social group. Many works in this approach covered " society" in an abstract and explicitly incorrect way, for example, assuming

that individuals are distributed on a regular grid and interact only with their nearest neighbors.

As the research community grew, the models became more elaborate. In particular, attempts were made to include enhancements to the models, which would allow them to correspond to more complex social phenomena, in particular to describe polarization and extremism. Such extensions included adding agents' roles and behaviors (e. g., contrarian stances, inflexibility, leadership, and extremism), considering the opinion space as multidimensional (and, by including relationships between opinions on related issues, analyzing the opinion dynamics within such belief landscapes), or including noninformational aspects of processes driving individual opinion changes, for example, agents' emotions.

The last phase—currently at a nascent stage—may be called " post-truth" opinion modeling. It is driven, on one side, with increasing ability to gather data on opinions expressed in social media (thus allowing monitoring individuals' and social group opinions on a much more detailed level than opinion polls, consumer data, or election results), and on the other, by an increased realization that much of the influences shaping public (and individual) opinions on various matters are driven by fake news, manipulation, algorithmic filtering, and various forms in human biases.

In this section, we will list the main characteristics of the three modeling generations, in particular noting the challenges faced by them. While we try to cover most of the directions taken by the research community in the past

thirty or so years, this work is not a complete review of the field. For a recent, comprehensive review, the reader is encouraged to consult [3].

2. 1 Physics of Unphysical Systems

As noted above, the first generation of the opinion change models have " borrowed" many notions from physics, in particular from physics of magnetism. This included describing individual opinions through a one-dimensional parameter, which could be continuous or discrete. In the latter case, the opinions attributed to people had, by design, a lot in common with atomic spins. The external influence was introduced through parameters corresponding to external magnetic field, and the volatility of individual opinions was connected to " social temperature." The analogy between human behavior and atomic spins has led to coining of a notion of spinson (spin-person, Ref. 4). The topology of social contacts was typically one of three variants: fully connected network, in which any person could communicate with any other person; a regular grid (typically two-dimensional square lattice, the easiest one for graphical presentation of results), where the agents could interact with their neighbors (either only the closest neighborhood or a suitably chosen extended one); a random network of a given density of links. Somewhat later on, the popularity of " network science" has led to models based on more complex interaction networks, including the scale-free topology, found to be present in many social systems.

The last component of the models was the interaction mechanism—describing the influence of one agent on others. This could take the form of one-to-one interactions (one agent trying to convince the other, or two

agents sharing a discussion and trying to convince each other, e. g., the voter model [5 – 8] and the bounded confidence model [9 – 12]) or many-to-one " group influence" models (where a suitably averaged opinion of a group " surrounding" an agent influenced its opinion, e. g., Galam models [13 – 17], the social impact model [18 – 22], or the Hegselmann–Krause model [23]). A rather complete review if this generation of opinion models is presented in Ref. 24 . The specific conditions required for an agent to keep/change its opinion varied between the models, leading to different process dynamics. There was, however, an assumption which was used in most of the early models: an interaction between two agents with differing opinions would diminish the difference or, at worst, leave it unchanged. This basic individual dynamics, in the absence of additional constraints (e. g., fixing certain agent's opinions or providing opinion change mechanisms, other than interactions between agents), should eventually lead to a convergence of opinions. In fact, many of the early works considered the dynamics of the appearance of such consensus, despite the rarity of such consensus in real societies. A review of the modeling field [25] noted that only a few opinion dynamics models made any reference to actual social systems.

The early works were later developed to include many new mechanisms, inhibiting this trivial path to consensus. Among the most widely recognized was the bounded confidence model, which assumed that the opinions of the two interacting agents could get closer to each other only if the initial difference was suitably small, below the tolerance threshold. Depending on the tolerance, this could lead to consensus (if the tolerance was high) or a

stable coexistence of two or more groups, within which the opinions would converge to different values. The bounded confidence model itself has been developed to cover many variants, some of them including repulsion of suitably different opinions, considerations of multidimensional opinion spaces, presence of minority groups holding fast to their opinions, and many others.

Another concept contributing to a potential lack of consensus in a society was to consider coupling of the individual opinion changes with the evolution of the social contacts network. The agents might be " allowed" to cut the social links with other agents if the differences in opinions were too high and to replace them with links to more like-minded agents. The result of this process is a separation of the community into unconnected parts, which was compared to sociological notion of echo chambers and selective attention. When the two processes—opinion change and social network evolution—are considered together, the final configuration may be quite complex and interesting. The topic of co-evolution of opinions and social networks has been studied, for example, in Refs. 26 – 30 .

The first-generation models had brought to attention several issues, some of which are not satisfactorily resolved to this day. One of them is the choice of starting conditions. Physics of magnetism typically uses the high-temperature limit in which the atomic spins are random and uncorrelated. Lowering of temperature leads to ordering and eventual appearance of interesting, nonrandom phenomena. Most of the models of opinion dynamics have started from a similar random state. However, the assumption that social systems can be " prepared" in such fully randomized way is not

realistic. What is worse? Starting a simulation from some partially ordered state may lead to results specific to these conditions. A different initial state can lead to the final one that is grossly different from the previous case. So, how do we know if we are starting the model from the " right" conditions? Most models continue to be initiated randomly, hoping for some universality of the final configurations, but this universality is by no means assured. Some recent models analyze other ways in which randomness may influence the system behavior, for example (Ref. 31), consider the role of noise (defined as randomness in adopting opinions) in consensus formation.

The second problem of the first-generation models is the mapping of the model time-ticks to real-world time. First, while the spin flops of single atoms can be considered instantaneous, when compared, for example, to the timescale of the sample temperature changes, the time it takes a person to process information and respond to an external influence by changing the previously held opinion may not only be quite long, but can also significantly vary between people and circumstances. Second, the same or even greater diversity may be present in the frequency of contacts. The events influencing our opinions may be separated by seconds (if we consider, e. g., a conversation as composed of separate messages) or by weeks or months, or, in the case of political campaigns preceding elections, even years.

In contrast, most ABMs measure time in simulation steps (or simulation steps per agent). And the scale of the simulations may span vast ranges: from hundreds of steps per agent to tens of thousands or even longer times. In many cases, the research focuses on reaching a final stable state (or at least some semi-stable one) of the system, which means that agents may have

extremely large number of " occasions" to flip their opinions. While such picture poses no problem in studies of magnetism, it seems rather distant from the intuitions and observations of human behavior. Mapping the simulation time to the real world is made even more difficult if we include the problems of the initial conditions. The assumption that we should focus on stable states is by no means necessary. It may well be that most of the social situations are transients, and the perception of equilibrium and stability comes from inadequate observation windows.

In some cases (e. g., reactions to significant one-time events, such as natural disasters or acts of terror), we can monitor the characteristic times of social behavior (e. g., the surges in charity donation rates or political reactions). This could provide an indication of the correspondence between models and reality, if the models exhibit similar large-scale time dependencies. In other cases, we can hope for even closer correspondence, for example, when the communication is mediated by social networking tools such as Twitter or Facebook or Internet discussions. If we treat each such message as corresponding to a single simulation event, we could test the model capability to mimic global dynamics using " realistic" numbers of time-ticks.

While the issues related to the initial conditions and to the understanding of timescales of changes observed in the simulations have been known for years, they are far from resolved and are still subject of active research, for example, Ref. 32 .

The first-generation models of opinion dynamics are still present in the current literature, despite their limited application to social sciences. One of

the reasons is that the flexibility of the topology of the interaction network, coupled with possible ways of describing the individual opinions (starting from the simplest, binary " spin-like" opinions, through discrete multi-valued opinions, to continuous versions and multidimensional ones), creates a large number of " unphysical systems" which can be studied with the focus on typically physical characteristics. It should be no surprise that most of such articles are published in journals traditionally devoted to pure physics.

2. 2 Enhanced Models

The second generation of models resulted from the expansion of interest of the researchers to the actual social systems and their properties. The differences from simple analogy to magnetism included such highly relevant phenomena as the role of leaders and media; the appearance, long-term stability, and influence of extremism; persistence of minorities based on opinions and worldviews; and the global resurgence of increasing polarization.

The last issue has been a true puzzle for the models as far back as the earliest attempts to provide mathematical framework to opinion dynamics [33]. That the problem remained valid after half a century may be confirmed by Ref. 34 , which put the question *how groups become polarized or how two groups can become more and more different?* in their list of important unsolved problems of sociology. Similar sentiment was voiced in Ref. 3 . Due to the social importance of the persistent conflicts and increasing division spanning all levels of democratic societies on many issues, the second generation of opinion models often focused on the issue.

One way of achieving polarization is through assuming an existence of special classes of agents (inflexibles, zealots, or extremists: [35 – 45]). If inflexible agents represent opinions at the extreme ends of the allowed spectrum, they may drive the opinion dynamics and " pull" the moderates toward them, creating a strongly polarized society. In fact one does not have to include any special class of inflexibles into the model, for example, within the bounded confidence framework, all that is required is a condition that the opinions close to the extreme ends of the opinion space are associated with decreasing tolerances, a rather intuitive observation. In such situation, most of the agents not only move to the extreme opinions but also at the same time, they become inflexibles themselves [46].

There are other ways to arrive at a polarized final state, for example, considering multiple topics, some of which may be more " important" to the agents, where disagreement on the more important opinion may drive polarization in the less important one [47 – 49]. The " multidimensional" models may become quite complex, as demonstrated by Ref. 50 ; combining internal dimensions characterizing the agents (education level and socioeconomic) and topic dimensions (relative to financial and social issues). The number of parameters in such model is usually quite large and allows similar systems to evolve either toward consensus or diversity of opinions.

Encounters between people in which they discuss their opinions often involve the arguments that are used in support of these opinions. When two persons share similar views, they can intensify their opinions by providing each other with new ways of supporting them, becoming more extreme in

result. This approach, dubbed persuasive argument theory, has been used in several polarization models, for example, Refs. 51 – 55 .

Still another approach is to include repulsive reactions to opinions similar to one's own, via nonconformism and contrariness [56 – 63]. There are also models stressing dissimilarity effects in which the agents do not only move their opinions closer to each other, if their opinions are relatively similar, but also split further apart, if the initial difference is greater than a certain threshold (e. g., Refs. 64 , 65 . Such " repulsive force" models have some justification from psychology, in particular in the studies of the " backfire" effects [66 – 70]. When confronted with information contrary to the current beliefs, for example, when two agents holding opposing opinions meet, instead of changing their opinions toward some averaged value, they may move in opposite directions.

At the same time, in considering the effectiveness of the backfire effect and its contribution to continued polarization, one should take into account the selective attention and confirmation bias [71 – 74], which severely limit the frequency and importance of encounters with opposite views. If we avoid contact with people or media presenting opposite views, we might be less angered by them, and the effects of repulsion would be diminished. The bias toward interactions with the agents who are sufficiently close to our opinions results in splitting of the society into weakly or noninteracting groups. Such effects were modeled in the original bounded confidence approaches [9 – 12], as well as in the network-based models assuming dynamic nature of the social links [30 , 75 , 76]. Thus, polarization understood as division of a society into separable groups holding nonoverlapping opinions can be

explained by the above mechanisms. On the other hand, without repulsive forces, it is difficult for the models to reproduce the evolution toward more and more extreme opinions, often associated with polarization. A recent survey of advances in opinion dynamics [77] lists several " milestone" extensions to the classical models, most of which were already mentioned here: stubbornness, presence of bias, manipulation of opinion (in the meaning of media influence), presence of repulsive behaviors, interrelations between multiple topics, and the difference between expressed and private opinions. An interesting direction, pointed out by Noorazar, which we have not discussed so far, is the introduction of the evolution of power in a social network.

As one can see, there are multiple solutions proposed to resolve the problem of polarization and, in general, the complexity of opinion dynamics. They differ not only in the " mechanics" of the models but also in their reference to psychological and social theories used as backgrounds. This is not surprising: the wide variety of models reflects an even greater variety of explanations proposed in social sciences, often connected in complex ways or even mutually contradictory.

The second generation of models' research interests was not limited to studies of polarization. Another interesting direction, initiated quite early, was the study of the role of leaders. Initial work in this direction started quite early [21 , 22 , 78 – 81]. It is somewhat surprising that the topic has not attracted a greater attention recently (there are relatively few articles on the subject, e. g., Refs. 82 – 86 in the light of the modern day phenomenon of " influencers." This can be explained partially by the focus on the scale-free

network topology and the natural assignment of the " leader" role to the highly connected network hubs. Such emphasis on the network characteristics of high-influence agents/people does not explain the origin of new opinion leaders and extremely interesting time dependencies of their rise of popularity and its waning. With the current availability of the Internet-based data concerning various areas in which the influencers operate (culture, fashion, lifestyle, politics, etc.) the leadership models certainly deserve even more attention.

Another direction of development of the opinion dynamics models was brought by combining them with studies of sentiment (or emotion) in interpersonal communications, especially those conducted via electronic means. The development of effective sentiment analysis software, such as SentiStrength [87 , 88] has allowed to classify medium- and large-scale datasets of Internet discussions and social media activities. It has been found that emotions not only " accompany" exchanges of opinions but also, in many cases, determine the outcomes of such exchanges and shape the communication processes. Starting from these observations, ABMs including emotions as parameters allowed to reproduce a wide range of characteristic features of communications on a wide range of topics [89 – 94]. The success of combining information and emotions in a single framework has led to further development of ABMs devoted to a broader range of issues. For example, the emotion/information/opinion (EIO) model [95 , 96] used nonlinear dynamics of the interplay between emotions and information received to shape individual opinions. The model has allowed to predict, through fully abstract model, the results of Polish Parliamentary elections of

2015 with 3% accuracy [97]. The interest in modeling the role of emotions in various opinion dynamics contexts is very promising [98 – 101].

2. 3 Post-Truth World Modeling Challenges

Despite the advances offered by the second generation of models (e. g., the capacity to mimic not only consensus but also conflicts and disagreements), there is still an enormous gap between the wealth of actual human behaviors and the limited range of actions and characteristics included in the agent-based models.

In some cases, these limitations touch the very foundations of the early model paradigm. For example, the information transmitted in person-to-person contacts or received from media or observations might be inaccurate, manipulated, or false. The origin of the difference between the communicated values and the true state (e. g., the true opinion of the person we talk with or an accurate description of the matter related by media) might be intentional or not. Conscious lies might be driven by an effort to achieve certain goals, to defend or preserve one's position in a group or society. But false information might also result from self-delusion or self-deception. Moreover, purposeful misrepresentation is not necessarily " evil"— manipulation of information is used in many situations " for the benefit of all interested," including the recipient (who would then make " correct" decisions based on erroneous data). Inclusion of misrepresentation and manipulation into the modeling framework is extremely difficult, as it depends on our ability to assign goals and motivations to the agents. There is very little we know about the statistical distribution of such characteristics,

so creating sensible initial conditions, especially for large-scale simulations, would be next to impossible.

Even more deeply, the influence of the communication on particular person opinion may be (and often is) dependent on one or more cognitive biases, triggered by the communication content, form, or even by external circumstances. Psychology research recognizes over a hundred different (named) types of biases [102 – 105], resulting from factors such as information overload, lack of meaning, the need to act fast, and selective information processing and memorization. Which biases dominate for a particular person in a given situation is far from obvious and may depend on the personal history or on circumstances that are typically absent in the modeling framework.

Some of these biases are relatively simple to include in small-scale simulations [106]. The general integration of the biased processing into the ABM approach is still missing, although there are recent works tackling parts of the problem, for example [107], which shows how three levels filtering the information (individual, social, and technological) may be incorporated into the modeling framework.

In Section 5, we will attempt to list some of the components and ideas for the future model generation allowing to cover the post-truth world. However, before we move to these suggestions, it is worthwhile to consider the general goals of modeling (Section 3), as well as some most pressing examples of situations where ABMs would have significant value (Section 4).

3 Why Model?

Perhaps to answer in which direction the opinion dynamics models should develop, we should ask a more fundamental one, concerning their goals. The use of models, which rely on selected features of the observed phenomena and provide tools to describe and predict the behavior of the studied system and to compare it with the observations, is the cornerstone of modern science. Mathematical models have allowed physics to go from historical descriptions of observations to an operational understanding of the world around us.

Observing regularities of physical and (especially) astronomical systems prompted model creation. In return, studying the model behavior allowed to predict future events and behaviors (such as eclipses or trajectories of a projectile). Even in cases where early models were superseded by completely different and more accurate ones (one could refer here to the transition from Ptolemaic epicycles to Newton's laws of motion applied to planets and then to Einstein's general relativity), the very capacity of constructing such models was instrumental in making physics (and later, chemistry and biology) so successful.

Understanding human behavioral patterns and social systems has never reached a similar accuracy and reproducibility. The dream of " social physics" is still unfulfilled. The matter is simply too complex, too difficult. In many cases, the observations are reduced to anecdotal science, or to grossly simplified generalizations. The efforts to create a quantitative framework in social sciences are often limited to statistical tools. And because it is extremely difficult to control all the conditions of experiments or

observations, the reproducibility in sociology and psychology is quite low, which has recently been recognized as an important problem.

The agent-based model approach may provide a partial solution to these limitations of psychological and especially social studies. Reference 108 asked the title question and suggested some meaningful reasons, other than prediction. These included goals related to deeper understanding of the situation, such as explanation, understanding of core dynamics, or illuminating uncertainties. Other goals bridged the gap between models and empirical studies, with models providing guidance for data collection, showing inconsistencies between theories and data and indicating new questions that could move empirical sociology and psychology forward. Epstein's reluctance to promote prediction as the main goal of the models has prompted an important discussion [109 , 110] which focused on the meaning of the notion of prediction. In the realm of social sciences, where events are often unique or only qualitatively similar, prediction and comparison with empirical observations play a less dominant role than in physics or astronomy. But if we accept a more general definition of predictions—generation of expectations of the studied system behaviors, the links between explanation and prediction become much stronger.

The epistemological status of simulations has also been discussed by Ref. 111 . It suggested answer to the " why model?" question being " to develop a mental model" which might seem circular. A model to develop a model? How is the computer model (simulation) different from the mental one? One possible answer is that by using computers, we can " run" multiple scenarios, following dynamics too complex or too computationally intensive

for our thinking, looking for statistical regularities or tracing the important development paths. The results of the simulation may be quite unexpected, even if we know the conditions embedded in the model. And this very " embedding," needed to create a computer simulation, requires the researchers to internalize the relative importance of the processes and their limitations, providing a deeper understanding of the studied process. Discrepancies between model results and observations point out that some of the assumptions might be wrong.

Shifting our attention back to physical sciences, this focus on unexpected differences between theoretical models and observations is crucial. It might signify deficiency in the model (most often) or inaccuracies or errors in observations or experiments (less common case). The use of mathematical models has not only brought precise predictions of various phenomena but also understanding of the underlying reality. The derivation of Newtonian gravity from Kepler's laws, or quantum mechanics from discrepancies of atomic physics by the beginning of 20th century clearly show that models can drive explanation.

Have these goals of explanation and understanding been really achieved in opinion dynamics research? Despite the increasing complexity of the models and growing volume of publications, it is hard to provide examples of which models have suggested a solution to open questions in social sciences. Even in the cases where simulations correspond to realistic cases, the modeling community is often happy with achieving a qualitative reproduction of observed phenomena (more about this in the following section).

There is, however, another way in which computers are used in science, which has risen in prominence since the works quoted above. The developments in readily applicable artificial intelligence tools (also referred to as machine learning, deep learning, or deep neural networks) have resulted in a wave of research using these tools to " solve" problems in many disciplines. Articles with titles or abstracts containing the phrase " use of deep learning in …" (or similar) and claiming breakthrough advances may be found in such " hard sciences" as physics, astronomy, and (bio)chemistry. Sometimes they provide incredible speedups over traditional methods, based on direct application of known dynamics laws (see, e. g., Ref. 112 . But these undoubted achievements often lead to questions regarding the potential of deep neural models to increase our *understanding* of the phenomena. The AI models are " trained" by requiring " similarity" with some previously analyzed sets: configurations leading to " wrong" results are culled out. Would the final neural network with optimized parameters be capable of finding new phenomena, or do we introduce a strong bias into our tools? Another issue is connected with the fact that the AI models are, for all practical purposes, black boxes. We do not know which parameters are " important" and which are secondary; we do not gain knowledge why the studied system evolves this or that way. In cosmology or materials physics, researchers have the " backup" strategy of existing fundamental models, but in the case of modeling of social phenomena, this is not the case.

While the deep learning models are not prominent (yet) in the scientific studies of social behaviors, there are indications that they are used in commercial environments, such as banking (for the evaluation of credit

status), Internet media filtering, and advertising or in government activities (e. g., predicting criminal behavior). The biases present in these models, resulting from the biased training sets, cause well-justified protests against the use of the AI tools in sensitive areas.

3. 1 Linking Modeling With Psychological Theories and Empirical Sociology

A deepened relationship between traditional social sciences and agent-based models can be mutually beneficial. As already noted, models may add the capability of qualitative (and, in the future, quantitative) predictions of social systems behaviors. When properly constructed and optimized, the ABMs may also increase our levels of insights into human behaviors, especially with respect to large groups, fact noted outside the ABM community [113]. While the predictive capabilities may have a dark side (see next section), the improved understanding of the actual mechanisms, beyond single person or small group experiments, can not only move science forward but also allow us to respond better to societal needs.

Models may also prove to be valuable in the " measurement revolution" brought to social sciences by " big data" analyses [114]. ABMs may become the link between huge sets of digital traces we leave online and the theoretical concepts and debates in psychology and sociology.

On the other hand, the modeling community would greatly benefit from a development of a common language with social sciences, from an improved ability to ask theoretical and empirical questions in a way that provides input to models. This input may take the form of initial data or suggestions

regarding the intrinsic model dynamics (of the agents or of the influence processes).

Unfortunately, such multidisciplinary approach is currently quite rare. The methodologies used by the social scientists (where quantitative analyses focus on finding which theoretical parameters have the highest statistical significance) and the modeling community (where focus is on model evolution and scenario outcomes as functions of model parameters) need to be connected. Moreover, the two groups of researchers typically publish in and read separate sets of research journals, which inhibits communication and creation of new ideas. Similar ethical issues may be present in the case of opinion dynamics models (see Section 3. 2).

3. 1. 1 Can Simple Opinion Dynamics Models Provide Meaningful Results?

Simple, even simplistic, models, which clearly lack the depth to match the real-world systems can, nevertheless, be useful. This situation happens when we can clearly distinguish the dominant processes and when we are interested in studying the dynamics and conditions connected to them. Of course, to be useful, the models have to offer something beyond crude analogies (such as " consensus = full magnetization").

A possible goal would be to focus the model on a single measurable aspect and try to find the driving forces. An example—so far less studied than it deserves—may be provided by modeling the relaxation of opinions resulting from high impact, emotionally loaded spikes; in other words: how quickly we forget what seems to be the most important issue at the moment, and which individual and social mechanisms drive this process. Is there any threshold

above which the impact of such driving event becomes (semi-)permanent? Is the decrease of importance mainly due to a gradual memory fading, or to competition with more recent, and equally " important" events? How do individual memory effects " translate" into collective memory? What processes govern habituation to events spanning long periods of time?

As the above example shows, the " relaxation" model does not need to be very complicated. Moreover, because of the dominance of social networks in large-scale communications, we may expect that there would be ample empirical data to compare the models with reality. There are excellent examples of recent works exploring this direction [115 – 117].

3. 1. 2 How Can We Use Complex Models Without Getting Lost in Details?

While simple models' weaknesses lie in the disregard of some aspects of the situation, the complex models are mired in our lack of understanding of the interactions of multiple dynamical mechanisms and scarce information about the initial conditions which should be used in the models. In large part, this may be traced to the state of knowledge provided by sociology and psychology. Social sciences offer multiple, often contradictory, " explanations" of our behaviors, and it is quite possible to build ABMs, using these diverse explanations as the underlying behavioral mechanisms.

This multiplicity of explanations can be seen already for simple models. The examples below are taken from the current author's limited forays into the field of modeling opinion polarization. They show, how different, often separate mechanisms can lead to similar results.

The lack of capacity to achieve a common viewpoint can be attributed, for example, to the interplay between individual opinion changes and the adjustments in the social network, due to avoidance of contacts with agents holding opposite opinions [30]. In such a model, the appearance of polarization (understood as stable division of opinions into two groups) depends on the ratio between freely modifiable social links and those which are fixed (e. g., family or workplace relationships).

Another way of " explaining" polarization may be via taking into account emotions and emotional commitment of the agents [95 , 96]. The nonlinearity of interactions between received information and emotional state has been sufficient not only to describe polarization and minority persistence but also to predict results of actual elections in Poland [97], using (in contrast to the previous example) a simplistic, fixed square lattice social network topology.

Extreme polarization (growth of the separation of opinions of opposing groups) has often been explained through inclusion of special class of agents (extremists, inflexibles, etc.). However, as shown in Ref. 46 , such inclusion is not necessary. What suffices is a simple mechanism linking the more extreme views (in a continuum space) with a decreased tolerance for other positions.

The last example (Ref. 106) focuses on the issue of biases in information processing by individual agents, including phenomena such as confirmation bias, motivated reasoning, and memory effects. Complex " biased" agents can polarize, even in a radically simplistic social environment.

All the above models have corresponding psychological or sociological foundations, either theoretical or empirical. Because all of them point into the same " direction" of a polarized society, their combination in a single complex model would (probably) raise only the question of the dominant mechanism: what drives the polarization the fastest in a given environment. This issue is, in fact, present in many social sciences articles.

The situation becomes more complex when the competing mechanisms point in different, often contradictory, directions. Assigning correct weights to various elements of the model dynamics may result in completely different results. Consider, for example, a potentially very important application of opinion dynamics models to the case of determination of which social activities and policies could lead to a decrease in the currently observed high levels of polarization and intergroup aggression. The individual- and group-based mechanisms combine in ways that are far from linear. This nonlinearity seriously impacts the usability of ABMs in such cases —unless we can come up with good controls, or novel simplifications preserving and focusing on the core part.

Complex models face one additional obstacle: they require much more initial data to run. Starting conditions describe both the individual and social interaction levels. Where to get such data from? How to assign the initial conditions, in a way that would not lead to spurious effects (as is the case of random initial conditions)? Are there correlations in the initial data that have to be taken into account? Such questions cannot be answered by the modeling community alone; we have to link more closely with social sciences practitioners. But as sociology and psychology rely (in their " hard" core) on

statistical descriptions, even a close cooperation between the modelers and social scientists may miss the effects of individuals or small groups, anything not captured by statistically expressed empirical descriptions.

So, are the efforts to develop more and more complex models misguided? Would we get lost in an exponentially growing number of parameters and their interactions? Perhaps part of the answer would lie in borrowing methods and approaches of disciplines faced with equally complex problems. Examples of such fields may be provided by climate modeling or by the challenges found in multiscale models of complex materials. Both these fields need to include phenomena occurring at various time and size scales. In both, it is simply impossible to use the most accurate dynamics at the lowest scale for the whole modeled system. The response, developed in both disciplines, was to create a hierarchy of scales and to create tools for transitions between the levels of such hierarchies. In fact, these transitions, defining what characteristics must be preserved in detail, and which can be simplified (e. g., averaged or considered locally stable) are the most challenging in achieving the high-quality models.

Of course, it is possible to create and investigate complex models with the goal of studying the models themselves. There might be some very interesting consequences of the nonlinear interactions between various mechanisms and abstract initial conditions—but I leave the reader to decide whether such approach, decoupled from empirical focus, is important enough.

Do these obstacles mean that there is no point in developing complex models? Not necessarily. First, there might be social situations in which we do know enough to construct a model that is sufficiently realistic in its details and in the knowledge of initial conditions. This could happen in two extreme circumstances: either for small, highly controlled groups, with agents designed to mimic specific individuals and their behaviors or in highly homogenized, large groups, such as crowds united by strong emotions (fear, anxiety, and joy), where the dominant agent states may be sufficiently uniform.

Another case where complex models might be useful is when a combination of several mechanisms creates a positive feedback loop, resulting in a system dynamics that is largely insensitive to the details of the starting setup. The value of the model is then in illuminating the way in which these different, but coupled, processes (each of which alone might not be sufficiently potent) reinforce each other. Models of this kind might be important in practical terms, helping to understand how radicalization of views may encompass large parts of societies, leading to undesirable behaviors. One could also use them to look for processes and mechanisms which could disrupt the feedback loop and inhibit crowd mentality.

3. 2 Ethical Dilemmas

One of the lessons of the recent decade is that the availability of increasingly detailed data on our behaviors, and the tools to analyze it open the way not only to understand social behaviors but also to monitor (often in real time) and manipulate behaviors, both on an individual level and for social groups. The tools initially focused on the data stored in the records of our Internet

activities (some of which we provide voluntarily, some of which are collected without our knowledge, thought with formal approval [1]). Today, the social networks and Internet activity datasets are (may be) coupled with records from CCTV monitoring, face and speech recognition algorithms, mobile phone data, including geographic positioning, etc. The advances in AI technologies further improve the whole system capacity to record and analyze our patterns of behavior.

Compared with the wealth of such data and the largely unknown capacities of algorithms and applications used by various government agencies and commercial companies, the research community has relatively limited access to sensitive data and much less freedom to experiment. The ethical guidelines and restrictions on academic research are quite stringent. Despite these constraints, it is possible to use ABMs to predict results of political campaigns and analyze " what-if" scenarios in politics with publicly available data, as shown, for example, by Ref. 97 . In this context, we suspect, with quite large probability of being right, that " powers that be" are not only using the digital traces we leave behind us to predict our behaviors but also funnel our attention and manipulate our perceptions. This manipulation is not limited to facts but covers also our information about what other people think and feel.

In the case of scientific research, serious, large-scale manipulation is, fortunately, out of question. The few exceptions, such as Ref. 118 , are actually proving the rule. Using truly personal, multifaceted data (combining expressed opinions and social activities with detailed personal information,

containing, e. g., the socioeconomic status, education, family, residence, and employment history) are mostly impossible. From one point of view, the ethics boards and committees do a great job, inhibiting misuses of power in the name of science. On the other hand, we do not really know what are the true capabilities of those entities (commercial, government, and illegal) with less stringent controls. All these ethical considerations have no simple solutions [119] but should be weighted by each researcher and team. What are our obligations and goals as scientists and as members of society? Can we add value beyond the creation of yet another abstract model? Can we describe and predict behaviors of real social systems? Should we? Without providing a general answer (impossible task), we would focus not on a couple of actual systems, focusing on their characteristics, which are quite difficult for the models to grasp.

4 Real-World Examples in Need of Good Models

To provide a better understanding and an illustration of the complexities of opinion dynamics modeling and the challenges posed by the real-world situations, we will now focus on two highly publicized examples. The first is the *anti-vaccination movement* . Many democratic countries have reported, in the past two decades, a significant drop in vaccination rates, sometimes considered as posing a real danger to societal health. These developments took place against clear and unified support for vaccination from the medical and scientific communities and from the governmental agencies. In many cases, the anti-vaccination activities grew, despite active administrative or legal countermeasures. While the empirical literature on the anti-vaccination

movement is quite rich, there is little effort to use ABMs to model it and understand the basic processes—a matter of high societal importance.

The second empirical topic worthy of modeling effort is the *growth of political polarization* (and the increased shift toward more extreme views at the expense of the political middle ground). As in the case of vaccination support, the political, social, and psychological literature devoted to extremism and polarization are quite rich. Here, there are many ABM-based efforts to explain the origins and mechanisms driving polarization, but we are still quite away from realistic models capable of more than qualitative similarities with actual cases. Below, we will describe some of the characteristic features of the two domains that pose challenges for the modeling approach.

4. 1 Anti-Vaccination Movement

The influence flows in the vaccination/anti-vaccination debate and the factors driving the individual decisions have some specific features, only a few of which are covered in the traditional and even in the extended families of models.

• Asymmetry in emotional content between the pro- and anti-vaccination influences. The latter ones often evoke the fear of health of one's own children—an extremely strong and motivating factor. In contrast, the pro-vaccination influencers (governmental, medical, and social actors) use either a nonemotional, rational arguments and language or appeals based on the interests of the society as a whole (such as threats to " herd immunity"). The pro-vaccination arguments in most cases miss the intimacy and urgency of

mother–child emotions, used effectively by the anti-vaccination movement. To model this difference in the communication modes and the associated reactions, one would have to use quite complex description of the influence process (including much more than simple comparisons of current opinions between agents) and much more complex agents as well, capable of contextually diverse reactions. These phenomena could, perhaps, be captured in emotion-based models [95 , 120], but even these models require many additional adjustments and modifications.

• Strong dependence of the anti-vaccination influence on the social status of the influencer, disconnected from the actual expertise and knowledge regarding the subject (the disproportionate influence of celebrities, such as actors/actresses). Interestingly, here, we can observe an asymmetry, too. The pro-vaccination stance of well-recognizable nonprofessionals appears to have less importance than a similar activity of anti-vaccination celebrities. This may be due to a lesser emotional appeal (see above) but also to the campaign to decrease the trust in experts by the anti-vaccination movement. This is achieved by portraying them as paid agents of pharmaceutical industry of governments. In contrast, the involvement of celebrities fighting vaccination is presented and perceived as coming from genuine concerns, quite often originating from the universally recognized and respected personal experiences and fears. As above, there are some models which incorporate a special role of high-influence agents (leaders), but the asymmetry has to be carefully introduced into them.

• Another complication is introduced by diversity of motivation and knowledge of the subject among the participants in the vaccination debate.

Compared to the general public, the anti-vaccination activists show much higher levels of *apparent* knowledge related to the issue. They are able to muster their arguments in a coherent way (despite the fact that most of these arguments are false or only partially true and misrepresented). An average person could be overwhelmed by case histories, scientific-sounding arguments, references to research (actual or invented), etc. Moreover, the activists have higher motivation and persistence in their influence efforts. In contrast, the general public (the majority that actually vaccinates their children) is much less prepared to argue for the vaccination and to counter the pseudoscientific jargon. And because of their lesser motivation (after all their kids are already protected), their involvement and levels of activity are smaller. Such differences could be included in an ABM model but would require great care to avoid triviality of results (where the model output simply follows directly from the input parametrization).

• Connected with the previous issue is the trustworthiness of the experts/champions of the two viewpoints. The role of the experts has been incorporated in the ABM previously [121], but the problem lies in the fact that the very status of an expert is rather complex. Psychological research shows that the agreement that a person is an expert in a given field depends on the cultural bias and opinion held by the evaluating person [122]. In addition to " raw credentials" (such as having a scientific degree or association with a research institution), the experts are judged trustworthy and influential on the basis of other factors. One is their alignment with currently held views (confirmation bias), considered in the ABM context, for example, by Ref. 123 . Another is the perceived motivation of the expert in

voicing their views. The anti-vaccination movement presents their experts as ethically motivated, caring, grassroots informal group. They are opposed by doctors or scientists who are—depending on the social environment—" representatives of the greedy pharmaceutical industry," " government lackeys," or " paid servants." Such negative branding may focus on either the government or " big pharma" (depending on which one is more reviled in a social group or country). The influence mechanisms focused on discrediting the voices of the opposing party are new and not considered in the opinion dynamics ABMs.

• At the individual person level, we deal also with many other forms of biased information processing. The attempts to include such biases in agent-based models are only just the beginning. Moreover, the biases may significantly differ not only between different people but also for a single person, depending on the circumstances. In the context of vaccination, the acceptance of various pro/con arguments may crucially depend on the presence of an illness in the family, friends circle, or immediate environment, or other transitory circumstances.

• The last element we mention here (noting that the list is by no means exhaustive) is the presence of systematic biases and manipulations of the influence streams. These may include efforts of governments and (through their pressure) censorship of social media providers. The models should not only have room for such systematic modifications of the influence patterns (such as silencing certain anti-vaccination voices) but also the countereffects, for example, the increased trustworthiness of these voices among the anti-vaccination community. This increase in strength of their

influence happens because of the attempts at their silencing. The pro-vaccination censorship is taken as the proof of the anti-vaccination arguments. From the point of view of the modeling, we need a much better understanding of the relative impacts of these manipulation/countermanipulation efforts.

4. 2 Political Polarization

The increases of political and social polarization and the related extremization of views may be observed throughout the world. They may spring from different origins: religious, ethnic, economic, or ideological; they may be also induced by skillful manipulation and propaganda, or be rooted in dormant fears and prejudices.

The importance of the polarization phenomena has been noticed by the modeling community: in the past decade, the number of articles tackling the problem has greatly increased, and, as noted above, several approaches have gained popularity. Among them, the most promising are models relaying on (and often combining) the role of a relatively small but active and devoted groups influencing much larger sections of societies and the potential lack of communication between the opposing groups, inhibiting the chances for reaching a less polarized societies.

The same increased attention to the polarization origins and mechanisms has been present in the social sciences. For example, Ref. 114 discuss the challenges in understanding the datasets obtained from " big data" mining. As they rightly state, the access to incredibly rich data from social networks, while alleviating many of the problems present in the traditional social

science methods, has not brought a deep understanding of the social and political phenomena. One of the indicated reasons, in our opinion the most important *one* , is the need to provide conceptual foundation in which the huge volumes of data could be understood. This actually is a huge change for models, which could provide a " test and understand" framework. Unfortunately, the problems present in the empirical description of political phenomena confirm the intuition that the process may be quite complex and depend on many factors, many of which have not been incorporated into the modeling frameworks.

• The first problem is related to potential biases present in the traditional social studies (such as surveys or self-reported characteristics and preferences). In many cases, the empirical data obtained this way may be skewed by attempts to present oneself in a way that is perceived as socially acceptable. In direct words: people would often manipulate their answers (consciously or unconsciously) to present themselves in a good light. In the context of political opinion models, this would correspond to a discrepancy between the expressed opinions and the actual ones (which may result in unexpected differences between survey/poll results and e. g., actual voting, where anonymity allows more " honest" actions). Moreover, the difference between the internal (actual) opinion and the publicly expressed one would not be context-free: a person can be more honest with " safe" social contacts (e. g., friends sharing the same political views), while showing more duplicity with strangers or influential people known to be of opposite opinion. Describing such self-censorship is an interesting challenge for ABMs.

• Most ABM models are focused on single, identifiable issues, expecting a clear opinion of an agent. In political practice, we observe already dominant, and still growing, importance of the political affiliation and submission to the " party line" at the cost of actual individual opinions on specific subject. To paraphrase: " my party, right or wrong." Thus, in the ABM context, we should consider factors such as an individual motivation (inhibiting or promoting changes of opinion) and the " courage" to oppose the party views or the varied importance of the specific issues for the partisan political viewpoint (what belongs to an immutable core, for which no dissent is allowed, and what is left " free" to individually decide).

• Historically, most models focused on the convergent, stable final states. This is a heritage from equilibrium statistical physics (and equilibrium thermodynamics). However, even a cursory observation of social phenomena (political transitions, wars, social scares, and fashions) proves that we are dealing with transitory, nonequilibrium phenomena. Examples might be provided by relatively fast responses to one-time events (such as terrorism acts, natural catastrophes, election campaigns, and reaction to results), or slightly slower intergenerational social changes or effects of evolving technologies (for an example of analysis of such change, see Ref. 124 . The dynamic nature of societies, where some processes run much faster than others, means that the models cannot rely on the assumption that a process starting from a random configuration and settling down to some stable or metastable state is a valid approach. Significant attention must be given to the initial state, as well as to the proper timing of various phenomena influencing individual agents' opinions and behaviors.

• The political processes today are largely influenced by overt and covert manipulation, such as fake news, trolling, algorithmic profiling, and manipulation of content which reaches a person. Some people recognize fake news, and others not. Some people troll, some react to trolling, and some follow the adage " do not feed the troll." And many trolls are not humans at all but are perceived as persons. Who displays these behaviors depends on many factors and is not always repeatable, even for the same person. The algorithmic biases are now effective down to a single person level: googling the same set of words results in completely different results for most of us. What we see on Facebook or Twitter or TikTok is just as different. Thus, even the assumption that a rational agents should converge on a single " truth" is not valid: if they are fed separate and on different information streams, their worldviews would be fundamentally different. Thus, to be more realistic, the ABMs must include descriptions of the manipulative practices—which is a problem because we know only very little about the actual mechanisms, scope, and techniques used in the manipulations.

5 Framework for a Post-Truth Models

5. 1 Communications

Below we present a selection of ideas, which could be used in models of opinion dynamics responding to the challenges shown by the real-world situations. The first group of challenges and ideas focuses on the communication processes influencing the opinion changes.

5. 1. 1 Influence Channels

Many models already take into account differences in topologies of interpersonal communication. The evolution of interagent networks from simple lattice geometries to complex, data-driven social networks, including networks that change dynamically in parallel to opinion changes is already well-established in the literature [30 , 65 , 75 , 76 , 125 – 128]. However, there is no consensus on the treatment of the mixture of effects of interpersonal contacts and exposure to media (both traditional unidirectional mass media: TV, radio, printed media, and participatory/Internet based media allowing some form of response). The two influence channels differ not only in their basic topologies but also in the nature of the communication (bidirectional vs unidirectional; one-to-one, one-to-many, many-to-many; and in person vs. print/screen-mediated). These differences may significantly impact the strength and durability of the influence.

Additional problem, from the point of view of the modelers, is to weigh in the relative frequencies of exposure between the interpersonal and media-driven communication (we would discuss the issue of strengths of influence in the next section). As the younger generations move ever larger parts of their life online, we might get some information from monitoring of peoples' activities.

2

In general, the lack of understanding of the relative strengths of various influence modes is not specific to the modeling community. The problem has not been resolved neither in psychology nor in sociology. The speed in which modes of communication and tools used to transmit information and emotions change, plus the ethical limitations associated with such detailed

studies, means that the research community understanding is often many steps behind the societal change.

5. 1. 2 Fake News, Trolls, Misinformation, and Manipulations

Fake news was always a part of our life. Rumors, accusations, and slanders are known in all societies. In the context of political campaigns, Ref. 129 provides a perfect historical background story, dating back over 150 years. Historically, there were limits on how false information could spread. The ease of creation and dissemination of false or misleading news has dramatically changed in the Internet age. Recent advances in deep fake techniques (such as the fake Nixon's speech announcing Apollo 11 astronauts' death) create a situation, where even well-prepared and skeptical person would not be able to tell the difference between manipulated and true records of events. The manipulation does not even have to be so technically advanced: the reliance on social media, implied trust in our social networks, or preferences and selective attention in traditional information providers allow simple manipulation through one-sided presentation of complex issues.

This situation spurred numerous serious studies, extending beyond anecdotal accounts. They cover the spreading of misinformation on social networks [130 – 135]; the ways in which it is processed [135 , 136]; emotional components of misinformation spread [137]; and effectiveness of countering misinformation [67 , 138].

The misinformation may originate at different places in a society (news/media organizations, groups, or individual persons). It may also be

distributed in various ways (traditional mass media, social networks, word-of-mouth, etc.). In addition, the reasons for spreading of fake news may differ enormously, from malicious intent to innocent forwarding due to honest concerns or even due to lack of recognition of the false nature of the message. For these reasons, the inclusion of fake news into the modeling paradigm is very challenging. Yet, due to its potentially very serious consequences, the fake news phenomenon would greatly benefit from additional understanding which could result from properly constructed ABMs.

A related, but separate, phenomenon, which may seriously impact opinion dynamics, is Internet trolling. The origins of trolls date back to early Usenet chatrooms, and since the 1980s, the trolling has evolved in multiple dimensions [139 , 140]. They now include elements such as emotional manipulation through provocation, incivility, abuse, but also concentrated attacks representing specific views (sometimes attributed to foreign meddling in political affairs, such as the " Kremlin trolls" [141 – 143]). Advances in AI and text processing have also resulted in appearance of automated (bot-based) trolling, in addition to human agents. Detection of troll networks (coordinated actions directed at persons, organizations, or ideas) is getting more and more challenging [144 – 146]. Moreover, in addition to purposeful trolling, where abusive and incivil behavior has manipulative goals, there are instances in which negative emotions are genuine [89 , 90 , 92 , 147 , 148]. Trolling—in its multiple varieties—is another challenge for modeling of opinion dynamics. At the same time, ABMs are probably uniquely positioned to study the effects of trolling and possible countermeasures, if we can find the right model framework.

5. 1. 3 Lies, Masks, and Acting

Most of the early opinion dynamics models assumed that agent communicate their opinions truthfully. As a result, an agent would " know" the true opinions of those who influence it (individual influencers, close neighbors, or societal averages). The change of the opinion of this agent was then based on the specific model's mechanism, but the mechanism took into account the correct information about other agents.

Yet we know very well, that in real life, we frequently do not know what others think. " Everybody lies." We try to present ourselves in a good light, we cover our mistakes and hide our true intentions behind masks and acting, or we think this is advantageous or safer. Outnumbered by the opposition, we often keep quiet or simulate agreement, keeping our thoughts hidden. Or the opposite: we want to appear as loyal members of the in-group and show more enthusiasm than we actually feel, toeing the party line.

The models should therefore be able to include these false perceptions of what other agents think. An early model introducing the distinction between the " true" internal state of the agent and its observable characteristics was the continuous opinion, discrete actions (CODA) model [56 , 149 , 150]. The model differentiated between more subtle (continuous) true opinions and cruder expressed ones (discrete choices). The differences were quite important—but still assumed that the externally observable actions correspond to the internal preferences of the agents. Reference 151 propose a model in which opinions are considered on two levels, public and private. It

combines the q-voter model and sociophysics with a complex four-dimensional model of social response, formulated by social psychologists.

The erroneous perceptions of other agents' opinions may result from two sources: internal biases (see next section) and false signals sent by others. In the case of interpersonal relationships, both depend on the motivations, goals, emotional states, and many other psychological and social factors. Moreover, there are situations where programmed, automated bots are perceived as real people. This results in a mix of complex psychological processes on the " receiver" side and goals and methods used by the bot owner (known or unknown).

In addition to the interpersonal communications, false information is present in the media, not just in the form of fake news (defined as pertaining to specific facts, events, people) but also as fake meta-news (defined as false information about what other people think about these specific facts, events, or other people; manipulated information about social norms and expectations).

5. 2 Internal Processes

To correspond to real human behaviors, the models of opinion formation and change should incorporate (or at least estimate the effects of) a large amount of nonintuitive phenomena, beyond the range of the simple equations used to describe individual actions and reactions. A good example is provided by the need to consider biased processing of information, especially the forms of biases which unconsciously distort the way we perceive the world or others' opinions or the influence of moods or individual

life histories/experiences on opinion preferences and associated activities (e. g., willingness or reluctance to spread and promote particular views). There are several domains which are largely neglected in the current modeling literature, yet which are known to be important for the individual opinion dynamics.

5. 2. 1 Belief Systems and Conceptual and Moral Foundations

Specific opinions, especially binary ones, often result from more general belief systems. These general foundations " allow" people to express definite views on issues on which they have very little knowledge.

There is growing evidence that at least a part of our moral preferences and conceptual framework may have mixed cultural and genetic components. Specific individual personality traits (e. g., big five categories: openness, conscientiousness, extraversion, agreeableness, and neuroticism, Ref. 152) ; moral foundations (e. g., dichotomies among care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, and sanctity/degradation, Refs. 153 – 155) drive individual behaviors and responses, including preference for certain belief systems.

The presence of such comprehensive mental systems, linking various domains and topics, has been barely touched in the modeling approach. It would be especially important for models of political opinion dynamics to consider what it means to be " a conservative" or " a liberal," and how these general stances could be included in agents' properties.

5. 2. 2 Narratives and Motivations

A closely related factor in the analysis is recognition of the fact that we often " perceive" the reality through narratives. Constructing stories, attributing motives (where motives are unknown or impossible to exist, as is the case of inanimate objects) allows us to " make sense" of certain situations [156 – 158]. But these stories often are in stark contrast with logic and may even cause us to disregard or twist factual information [130 , 159]. For example, the famous Linda effect, a staple of irrational reasoning [160] (where many people attribute more probability to a subset " Linda is a feminist bankteller" than to an encompassing set " Linda is a bankteller") may be simply explained not as an error in logical thinking but as a result of reliance on narrative perception: within a " story" the subjects were given, the " ending" of Linda becoming a feminist bankteller is much more probable.

What is important in the context of opinion dynamics modeling is that the narratives are extremely effective in directing individual opinions into desired positions [161 , 162], much more effective than " facts alone." However, incorporation of narratives into model construction is an extremely difficult task.

Another way in which we frequently distort information is motivated reasoning, phenomenon well known in social sciences [163 – 170]. It relies on emotions to create explanations, justifications, or decisions that fit the person's desires and goals, rather than those that accurately correspond to evidence. The influence of motivated reasoning may be the stronger the more intelligent and prepared is a person to twist the evidence to its

preferences and purposes. This creates a nontrivial dynamics of opinion change, including nonintuitive mechanisms, such as creation of effective " countermessages" by certain members of the society (one can remind here about the effort put into anti-vaccination " rebuttals" of medical information).

5. 2. 3 Trust and Involvement

Our reactions to the influences and information we receive from various sources (personal contacts and media alike) crucially depend on the level of trust we have for these sources. Trust is not simply a matter of holding the same views; in general, it is quite complex socially, even at the definition level [171 – 173].

In fact, trust has already been incorporated into opinion dynamics models (e. g., in Refs. 174 – 182) . However, even a cursory review of the literature on trust shows its complex nature. Examples may be differentiating levels of trust (individual, social, or abstract) [183]; the question whether trust matters more when the conflict of interest is small or large? [184]; roles of social preferences and adherence to social norms [185]; the question what factors undermine trust [186]; or the links between trust and belief and moral personality foundations [187]. As these examples show, it is rather hard to encode trust as a single parameter in an ABM, so there is a lot of room for improvements.

5. 2. 4 Individual Opinion Dynamics and Persistence

Almost every parent has had the experience of telling their children, with some measure of desperation, " How many times must I tell you that [xxx] is not good for you?" The question is extremely relevant to modeling of opinion

change, at a fundamental level. " Frequent change" solutions, such as assuming that every encounter results in opinion change (as often found in the spinson models), greatly underestimate the stability of our opinions. They lead to situations in which an agent might rapidly shift from one opinion to the opposite, depending on the information it receives (whom it interacts with). While this individual fluttering might slow down the overall process, it is not supported by psychology. We do not even have a proof that temporal averaging of frequent opinion changes would converge to a psychologically realistic average/final outcome.

The problem is that understanding the individual opinion stability/variability is not really well advanced in social sciences. Pertly, this comes from the difficulty of measuring the effect. The first question is whether opinion change is a result of separate events or a cumulative one? Can a single dominant cause of opinion change be named or is it only the straw that broke the camel's back?

The second issue is the persistence of the influence. A personal encounter or a media story might change a person's view on a specific topic. But is this change permanent? Especially when the new opinion stands in conflict with a broader system? For example, an emotionally loaded story of a tragic death of an immigrant child might result in acceptance of relaxing immigration laws. But would this change persist when the story vanishes from news-media headlines?

A few years ago, a study has appeared that addressed this very subject [188]. Not only the effect of single encounter has been found but also some

second-level effects were suggested. And the change touched very deeply entrenched subject of gay equality. The work included even some time measurements of the persistence of the effect. Unfortunately, the article methods were subsequently questioned [189] which led to the article retraction.

This retraction does not diminish the importance and validity of the research question. In a somewhat ironical twist of events, the same authors who questioned the work on gay equality, have published their own study [190] devoted to equally emotionally loaded topic of reducing transphobia. This article not only included measurements of effects of single canvassing but also studied temporal persistence and even resilience of the new opinion to opposing advertisements.

In light of social and environmental challenges, deep understanding of the effectiveness and durability of opinion-changing campaigns or actions is of great importance not only to the modeling community but also to social sciences as a whole. Unfortunately, all studies involving manipulation of peoples' opinions, especially in a way that might result in long-term changes is of high concern from the ethical point of view. The data on individual malleability of opinions and on temporal persistence are therefore vary scarce.

6 Conclusion: Back to Modeling Drawing Board

The preceding sections have listed numerous challenges facing the modeling approach to social opinion dynamics. Some technical solutions were already described, as were important topics, in which the ABM approach could be

very useful. These examples included the studies of the timing of rise and fall of popularity of certain issues/subjects or the influencer phenomenon (mentioned in Section 3. 1) or modeling of misinformation spread or trolling (discussed in Section 5. 1). Here, we would point out some general potential directions of the ABM methodology development.

The first goal should be to refocus on basics: individual behaviors and interactions. To resort to a physical analogy: while the simple Lennard-Jones interatomic potential can allow to describe some macroscopic properties of materials, it misses some important ones. To model them, more realistic descriptions of local interactions and atomic properties are necessary.

Such " nanoscale" knowledge can be achieved only in a close cooperation with social scientists. Combining experiments asking relevant questions in individual psychology and small group behaviors with models constructed specifically for the purpose of understanding such behaviors could lead to mutual benefits: for the social scientists and for the modeling community. Validation of models can happen via such experiments or via a thorough analysis of data. Such validation (and the associated generation of new concepts and ideas) requires much closer links between sociophysicists and representatives of psychology and sociology and would follow paths already taken by more mature fields such as biophysics.

The second direction could be to select cases in which at least some of the obfuscating factors listed in this work are absent or weak. In other words, to look for opinion dynamics is driven by only a few dominating factors and mechanisms, which could allow to " disentangle" the links between various

mechanisms and to create models which could improve our understanding of these processes.

One way of looking for such " clearer cases" is through big data analyses based on social networks. There are known limitations and problems associated with these sources, for example, the fact that they give access only to public personas, and not to the internal states of people. Especially interesting could be models of naturally occurring " Internet-based experiments" (to avoid the ethical issues connected to planned experimenting, such as described by Ref. 191 or Ref. 118 .

These developments could lead to new model paradigms, extending beyond the spinson analogy. Rather than attempting to create models which would describe all situations and societies, we could focus on models devoted to specific social circumstances. In the first case, we are faced with seemingly insurmountable obstacle of the model complexity necessary to contain all the psychological and social complexities. This would render the models to be too complex to learn anything from them. But by focusing on cases where a few specific conditions are crucial, the models might prove to be very useful. They could, in particular, help understand the nonlinearities of the interactions between competing mechanisms, something misses in purely statistical (as opposed to causal) social analyses.

The social opinion modeling community has not only grown enormously during the past 30 years but also begun to be recognized by social scientists. By combining our efforts, we could build a meaningful future not only for the

opinion modeling as a research field but also, hopefully, to provide value for our societies.

Author Contributions

PS: Conceived and wrote the paper.

Funding

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with the author.

Acknowledgments

Footnotes

1 The almost automatic agreement to requests for acceptance of Web site cookies and associated data processing has been the subject of a large number of jokes. And while, in principle, people recognize that there are " some" consequences of clicking the " I agree" button, almost all of us are guilty of doing it without any real consideration.

2 Which actually opens a very interesting topic of modeling dynamics of societies which are mixtures of different generations (to use U. S. terminology: boomers, Millenials, iGen), not only with differing preferences for social communication and information consumption but also with different socioeconomic status and position in societal hierarchies [124].

References

1. Schweitzer, F. Sociophysics. *Phys Today* . (2018). 71: 40–6. doi: 10. 1063/pt. 3. 3845.

2. Galam, S. *Sociophysics: a physicist's modeling of psycho-political phenomena* . Berlin, Germany: Springer (2012). 439 p.

3. Flache, A, Mäs, M, Feliciani, T, Chattoe-Brown, E, Deffuant, G, Huet, S, et al. . Models of social influence: towards the next frontiers. *J Artif Soc Soc Simulat* . (2017). 20: 2. doi: 10. 18564/jasss. 3521.

4. Nyczka, P, and Sznajd-Weron, K. Anticonformity or independence? – insights from statistical physics. *J Stat Phys* . (2013). 151: 174–202. doi: 10. 1007/s10955-013-0701-4.

5. Cox, J, and Griffeath, D. Diffusive clustering in the two dimensional voter model. *Ann Probab* . (1986). 14: 347–70. doi: 10. 1214/aop/1176992521.

6. Ben-Naim, E, Frachebourg, L, and Krapivsky, PL. Coarsening and persistence in the voter model. *Phys Rev E* . (1996). 53: 3078–87. doi: 10. 1103/physreve. 53. 3078.

7. Galam, S, Chopard, B, and Droz, M. Killer geometries in competing species dynamics. *Phys Stat Mech Appl* . (2002). 314: 256–63. doi: 10. 1016/s0378-4371(02)01178-0.

8. Castellano, C, Vilone, D, and Vespignani, A. Incomplete ordering of the voter model on small-world networks. *Europhys Lett* . (2003). 63: 153. doi: 10. 1209/epl/i2003-00490-0.

9. Deffuant, G, Neau, D, Amblard, F, and Weisbuch, G. Mixing beliefs among interacting agents. *Adv Complex Syst* . (2000). 3: 87–98. doi: 10. 1142/s0219525900000078.

10. Deffuant, G, Amblard, F, Weisbuch, G, and Faure, T. How can extremism prevail? A study based on the relative agreement interaction model. *J Artif Soc Soc Simulat* . (2002). 5.

11. Weisbuch, G. Bounded confidence and social networks. *Eur Phys J B Condens Matter* . (2004). 38: 339–43. doi: 10. 1140/epjb/e2004-00126-9.

12. Weisbuch, G, Deffuant, G, Amblard, F, and Nadal, J-P. Interacting agents and continuous opinions dynamics. In: Cowan, R, and Jonard, N, editors *Heterogenous agents, interactions and economic performance* . *Lecture notes in economics and mathematical systems* , Vol. 521. Berlin, Heidelberg: Springer (2003). p. 225–42.

13. Galam, S, Gefen, Y, and Shapir, Y. Sociophysics: a new approach of sociological collective behaviour. i. mean-behaviour description of a strike. *J Math Sociol* . (1982). 9: 1–13. doi: 10. 1080/0022250x. 1982. 9989929.

14. Galam, S. Majority rule, hierarchical structures, and democratic totalitarianism: a statistical approach. *J Math Psychol* . (1986). 30: 426–34. doi: 10. 1016/0022-2496(86)90019-2.

15. Galam, S, and Moscovici, S. Towards a theory of collective phenomena: consensus and attitude changes in groups. *Eur J Soc Psychol* . (1991). 21: 49–74. doi: 10. 1002/ejsp. 2420210105.

16. Galam, S. Rational group decision making. A random field Ising model at T= 0. *Phys Stat Mech Appl* . (1997). 238: 66–80. doi: 10. 1016/s0378-4371(96)00456-6.

17. Galam, S. Application of statistical physics to politics. *Phys Stat Mech Appl* . (1999). 274: 132–9. doi: 10. 1016/s0378-4371(99)00320-9.

18. Nowak, A, Szamrej, J, and Latané, B. From private attitude to public opinion: a dynamic theory of social impact. *Psychol Rev* . (1990). 97: 362–76. doi: 10. 1037/0033-295x. 97. 3. 362.

19. Nowak, A, and Lewenstein, M. Modeling social change with cellular automata. In: Hegselmann, R, Mueller, U, and Troitzsch, KG, editors *Modelling and simulation in the social sciences from a philosophy of science point of view* . Dordrecht, Netherlands: Kluver (1996). p. 249–85.

20. Holyst, J, Kacperski, K, and Schweitzer, F. Social impact models of opinion dynamics. *Ann Rev Comput. Phys.* . (2001). 9: 253–73. doi: 10. 1142/9789812811578_0005

21. Kacperski, K, and Holyst, J. Phase transitions as a persistent feature of groups with leaders in models of opinion formation. *Phys Stat Mech Appl* . (2000). 287: 631–43. doi: 10. 1016/s0378-4371(00)00398-8.

22. Kacperski, K, and Holyst, J. Opinion formation model with strong leader and external impact: a mean field approach. *Phys Stat Mech Appl* . (1999). 269: 511–26. doi: 10. 1016/s0378-4371(99)00174-0.

23. Hegselmann, R, and Krause, U. Opinion dynamics and bounded confidence models, analysis, and simulation. *J Artif Soc Soc Simul* . (2002). 5: 1–33.

24. Castellano, C, Fortunato, S, and Loreto, V. Statistical physics of social dynamics. *Rev Mod Phys* . (2009). 81: 591–646. doi: 10. 1103/revmodphys. 81. 591.

25. Sobkowicz, P. Modelling opinion formation with physics tools: call for closer link with reality. *J Artif Soc Soc Simulat* . (2009). 12: 11.

26. Holme, P, and Newman, MEJ. Nonequilibrium phase transition in the coevolution of networks and opinions. *Phys Rev E* . (2006). 74: 056108. doi: 10. 1103/physreve. 74. 056108.

27. Kimura, D, and Hayakawa, Y. Coevolutionary networks with homophily and heterophily. *Phys Rev E* . (2008). 78: 016103. doi: 10. 1103/physreve. 78. 016103.

28. Vazquez, F, Eguíluz, V, and San Miguel, M. Generic absorbing transition in coevolution dynamics. *Phys Rev Lett* . (2008). 100: 108702. doi: 10. 1103/PhysRevLett. 100. 108702

29. Nardini, C, Kozma, B, and Barrat, A. Who's talking first? consensus or lack thereof in coevolving opinion formation models. *Phys Rev Lett* . (2008). 100: 158701. doi: 10. 1103/physrevlett. 100. 158701.

30. Sobkowicz, P. Studies of opinion stability for small dynamic networks with opportunistic agents. *Int J Mod Phys C* . (2009b). 20: 1645–62. doi: 10. 1142/S0129183109014655

31. Kowalska-Styczeń, A, and Malarz, K. Noise induced unanimity and disorder in opinion formation. *PloS One* . (2020). 15: e0235313. doi: 10. 1371/journal. pone. 0235313.

32. Mukherjee, S, Biswas, S, and Sen, P. A long route to consensus: two stage coarsening in binary choice voting model. Preprint repository name [Preprint] (2020). Available from: https://arxiv. org/abs/2001. 11969 .

33. Abelson, RP. Mathematical models of the distribution of attitudes under controversy. In: Tucker, LR, editor *Contributions to mathematical psychology* . New York, NY: Holt, Reinehart and Winston, Inc. (1964). p. 142–60.

34. Bonacich, P, and Lu, P. *Introduction to mathematical sociology* . Princeton, NJ: Princeton University Press (2012). 240 p.

35. Galam, S, and Jacobs, F. The role of inflexible minorities in the breaking of democratic opinion dynamics. *Phys Stat Mech Appl* . (2007). 381: 366–76. doi: 10. 1016/j. physa. 2007. 03. 034.

36. Crokidakis, N, and de Oliveira, PMC. Inflexibility and independence: phase transitions in the majority-rule model. *Phys Rev E* . (2015). 92: 062122. doi: 10. 1103/physreve. 92. 062122.

37. Galam, S. Stubbornness as an unfortunate key to win a public debate: an illustration from sociophysics. *Mind Soc* . (2016). 15: 117–30. doi: 10. 1007/s11299-015-0175-y.

38. Verma, G, Swami, A, and Chan, K. The impact of competing zealots on opinion dynamics. *Phys Stat Mech Appl* . (2014). 395: 310–31. doi: 10. 1016/j. physa. 2013. 09. 045.

39. Sznajd-Weron, K, Tabiszewski, M, and Timpanaro, AM. Phase transition in the Sznajd model with independence. *Europhys Lett* . (2011). 96: 48002. doi: 10. 1209/0295-5075/96/48002.

40. Kashisaz, H, Hosseini, SS, and Darooneh, AH. The effect of zealots on the rate of consensus achievement in complex networks. *Phys Stat Mech Appl* . (2014). 402: 49–57. doi: 10. 1016/j. physa. 2014. 01. 055.

41. Mobilia, M. Nonlinear q-voter model with inflexible zealots. *Phys Rev E* . (2015). 92: 012803. doi: 10. 1103/physreve. 92. 012803.

42. Colaiori, F, and Castellano, C. Consensus versus persistence of disagreement in opinion formation: the role of zealots. *J Stat Mech Theor Exp* . (2016). 2016: 033401. doi: 10. 1088/1742-5468/2016/03/033401.

43. Fudolig, MID, and Esguerra, JPH. Analytic treatment of consensus achievement in the single-type zealotry voter model. *Phys Stat Mech Appl* . (2014). 413: 626–34. doi: 10. 1016/j. physa. 2014. 07. 033

44. Mobilia, M, Petersen, A, and Redner, S. On the role of zealotry in the voter model. *J Stat Mech Theor Exp* . (2007). 2007: P08029. doi: 10. 1088/1742-5468/2007/08/p08029.

45. Fotouhi, B, and Rabbat, MG. The effect of exogenous inputs and defiant agents on opinion dynamics with local and global interactions. *IEEE J Sel Top Signal Process* . (2013). 7: 347–57. doi: 10. 1109/jstsp. 2013. 2245629.

46. Sobkowicz, P. Extremism without extremists: deffuant model with emotions. *Front Phys* . (2015). 3: 17. doi: 10. 3389/fphy. 2015. 00017

47. Baldassarri, D, and Bearman, P. Dynamics of political polarization. *Am Socio Rev* . (2007). 72: 784. doi: 10. 1177/000312240707200507.

48. Huet, S, Deffuant, G, and Jager, W. A rejection mechanism in 2d bounded confidence provides more conformity. *Adv Complex Syst* . (2008). 11: 529–49. doi: 10. 1142/s0219525908001799.

49. Huet, S, and Deffuant, G. Openness leads to opinion stability and narrowness to volatility. *Adv Complex Syst* . (2010). 13: 405–23. doi: 10. 1142/s0219525910002633.

50. Ravandi, B, and Mili, F. Coherence and polarization in complex networks. *J Comput Soc Sci* . (2019). 2: 133–50. doi: 10. 1007/s42001-019-00036-w.

51. Mäs, M, and Flache, A. Differentiation without distancing. Explaining bi-polarization of opinions without negative influence. *PloS One* . (2013). 8: e74516. doi: 10. 1371/journal. pone. 0074516.

52. La Rocca, CE, Braunstein, LA, and Vazquez, F. The influence of persuasion in opinion formation and polarization. *Europhys Lett* . (2014). 106: 40004. doi: 10. 1209/0295-5075/106/40004.

53. Liu, Q, Zhao, J, and Wang, X. Multi-agent model of group polarisation with biased assimilation of arguments. *IET Control Theory Appl* . (2015). 9: 485–92. doi: 10. 1049/iet-cta. 2014. 0511.

54. Fu, G, and Zhang, W. Opinion formation and bi-polarization with biased assimilation and homophily. *Phys Stat Mech Appl* . (2016). 444: 700–12. doi: 10. 1016/j. physa. 2015. 10. 006

55. Pinasco, JP, Semeshenko, V, and Balenzuela, P. Modeling opinion dynamics: theoretical analysis and continuous approximation. *Chaos, Solit Fractals* . (2017). 98: 210–5. doi: 10. 1016/j. chaos. 2017. 03. 033.

56. Martins, ACR, and Kuba, CD. The importance of disagreeing: contrarians and extremism in the coda model. *Adv Complex Syst* . (2010). 13: 621–34. doi: 10. 1142/s0219525910002773.

57. Galam, S. Contrarian deterministic effects on opinion dynamics: " the hung elections scenario". *Phys Stat Mech Appl* . (2004). 333: 453–60. doi: 10. 1016/j. physa. 2003. 10. 041.

58. Masuda, N. Voter models with contrarian agents. Preprint repository name [Preprint] (2013). Available from: https://arxiv. org/abs/1306. 4401 .

59. Borghesi, C, and Galam, S. Chaotic, staggered, and polarized dynamics in opinion forming: the contrarian effect. *Phys Rev E* . (2006). 73: 066118. doi: 10. 1103/physreve. 73. 066118.

60. Galam, S. From 2000 Bush–Gore to 2006 Italian elections: voting at fifty-fifty and the contrarian effect. *Qual Quantity* . (2007). 41: 579–89. doi: 10. 1007/s11135-007-9072-8.

61. Lama, M, López, J, and Wio, H. Spontaneous emergence of contrarian-like behaviour in an opinion spreading model. *Europhys Lett* . (2005). 72: 851. doi: 10. 1209/epl/i2005-10299-3.

62. Sznajd-Weron, K, Szwabinski, J, and Weron, R. Is the person-situation debate important for agent-based modeling and vice-versa?. *PloS One* . (2014). 9: e112203. doi: 10. 1371/journal. pone. 0112203.

63. Jarman, M, Nowak, A, Borkowski, W, Serfass, D, Wong, A, and Vallacher, R. The critical few: anticonformists at the crossroads of minority opinion survival and collapse. *J Artif Soc Soc Simulat* . (2015). 18: 6. doi: 10. 18564/jasss. 2663.

64. Jager, W, and Amblard, F. Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. *Comput Math Organ Theor* . (2005). 10: 295–303. doi: 10. 1007/s10588-005-6282-2.

65. Macy, MW, Kitts, JA, Flache, A, and Benard, S. Polarization in dynamic networks: a hopfield model of emergent structure. In: Pattison, P, Carley, K, and Breiger, R, editors *Dynamic social network modeling and analysis* . Washington, D. C.: National Academies Press (2003). p. 162–73

66. Nyhan, B, and Reifler, J. When corrections fail: the persistence of political misperceptions. *Polit Behav* . (2010). 32: 303–30. doi: 10. 1007/s11109-010-9112-2.

67. Lewandowsky, S, Ecker, UKH, Seifert, CM, Schwarz, N, and Cook, J. Misinformation and its correction: continued influence and successful debiasing. *Psychol Sci Publ Interest* . (2012). 13: 106–31. doi: 10. 1177/1529100612451018

68. Zhou, J. Boomerangs versus javelins: how does polarization constrain communication on climate change? *Environ Polit* . (2016). 25: 788–811. doi: 10. 1080/09644016. 2016. 1166602

69. Kaid, LL. Political advertising. In *Handbook of political communication research* . Mahwah, NJ: Rutlege (2004). p. 155–202.

70. Sanna, LJ, Schwarz, N, and Stocker, SL. When debiasing backfires: accessible content and accessibility experiences in debiasing hindsight. *J Exp*

Psychol Learn Mem Cognit . (2002). 28: 497–502. doi: 10. 1037/0278-7393. 28. 3. 497. CrossRef Full Text |

71. Knobloch-Westerwick, S, Mothes, C, and Polavin, N. Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Commun Res* . (2020). 47: 104–24. doi: 10. 1177/0093650217719596.

72. Nickerson, R. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol* . (1998). 2: 175–220. doi: 10. 1037/1089-2680. 2. 2. 175.

73. Westerwick, A, Johnson, BK, and Knobloch-Westerwick, S. Confirmation biases in selective exposure to political online information: source bias vs. content bias. *Commun Monogr* . (2017). 84: 343–64. doi: 10. 1080/03637751. 2016. 1272761

74. Kahan, DM. The politically motivated reasoning paradigm, part 1: what politically motivated reasoning is and how to measure it (2016). Available from: .

75. Iñiguez, G, Kertész, J, Kaski, K, and Barrio, R. Opinion and community formation in coevolving networks. *Phys Rev E* . (2009). 80: 66119. doi: 10. 1103/physreve. 80. 066119.

76. Su, J, Liu, B, Li, Q, and Ma, H. Coevolution of opinions and directed adaptive networks in a social group. *J Artif Soc Soc Simulat* . (2014). 17: 4. doi: 10. 18564/jasss. 2424.

77. Noorazar, H. Recent advances in opinion propagation dynamics: a 2020 survey. *Eur Phys J Plus* . (2020). 135: 521. doi: 10. 1140/epjp/s13360-020-00541-2.

78. He, M, Xu, H, and Sun, Q. Opinion evolution on a scale-free network with leaders. *Int J Mod Phys C* . (2004). 15: 947–53. doi: 10. 1142/S012918310400639X

79. Boccara, N. Models of opinion formation: influence of opinion leaders. *Int J Mod Phys C* . (2008). 19: 93–109. doi: 10. 1142/s0129183108011954.

80. Düring, B, Markowich, P, Pietschmann, J, and Wolfram, M. Boltzmann and Fokker–Planck equations modelling opinion formation in the presence of strong leaders. *Proc Math Phys Eng Sci* . (2009). 465: 3687. doi: 10. 1098/rspa. 2009. 0239.

81. Sobkowicz, P. Effect of leader's strategy on opinion formation in networked societies with local interactions. *Int J Mod Phys C* . (2010). 21: 839–52. doi: 10. 1142/S0129183110015518

82. Van Eck, PS, Jager, W, and Leeflang, PSH. Opinion leaders' role in innovation diffusion: a simulation study. *J Prod Innovat Manag* . (2011). 28: 187–203. doi: 10. 1111/j. 1540-5885. 2011. 00791. x.

83. Jalili, M. Effects of leaders and social power on opinion formation in complex networks. *Simulation* . (2013). 89: 578–88. doi: 10. 1177/0037549712462621.

84. Hegselmann, R, and Krause, U. Opinion dynamics under the influence of radical groups, charismatic and leaders, and other constant signals: a simple unifying model. *Netw Heterogeneous Media* . (2015). 10: 477–509. doi: 10. 3934/nhm. 2015. 10. 477.

85. Yuan, X, and Crooks, A. From cyber space opinion leaders and the diffusion of anti-vaccine extremism to physical space disease outbreaks. In *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation* . Berlin, Germany: Springer (2017). p. 114–9.

86. Zhao, Y, Kou, G, Peng, Y, and Chen, Y. Understanding influence power of opinion leaders in e-commerce networks: an opinion dynamics theory perspective. *Inf Sci* . (2017). 426: 131–47. doi: 10. 1016/j. ins. 2017. 10. 031

87. Thelwall, M, Buckley, K, and Paltoglou, G. Sentiment strength detection for the social web. *J Am Soc Inf Sci Technol* . (2012). 63: 163–73. doi: 10. 1002/asi. 21662.

88. Thelwall, M, Buckley, K, Paltoglou, G, Cai, D, and Kappas, A. Sentiment strength detection in short informal text. *J Am Soc Inf Sci Technol* . (2010). 61: 2544–58. doi: 10. 1002/asi. 21416.

89. Sobkowicz, P, and Sobkowicz, A. Dynamics of hate based internet user networks. *Eur Phys J B* . (2010). 73: 633–43. doi: 10. 1140/epjb/e2010-00039-0

90. Chmiel, A, Sobkowicz, P, Sienkiewicz, J, Paltoglou, G, Buckley, K, Thelwall, M, et al. Negative emotions boost users activity at bbc forum. *Physica A* . (2011). 390: 2936–44. doi: 10. 1016/j. physa. 2011. 03. 040

91. Sobkowicz, P, and Sobkowicz, A. Properties of social network in an internet political discussion forum. *Adv Complex Syst* . (2012). 15: 1250062. doi: 10. 1142/S0219525912500622

92. Sobkowicz, P, and Sobkowicz, A. Two-year study of emotion and communication patterns in a highly polarized political discussion forum. *Soc Sci Comput Rev* . (2012). 30: 448–69. doi: 10. 1177/0894439312436512

93. Sobkowicz, P. Quantitative agent based model of user behavior in an internet discussion forum. *PloS One* . (2013). 8: e80524. doi: 10. 1371/journal. pone. 0080524

94. Sienkiewicz, J, Chmiel, A, Sobkowicz, P, and Holyst, JA. " How online emotions influence community life," in: Holyst, J, editor *Cyberemotions* . Berlin, Germany: Springer (2017). p. 159–85.

95. Sobkowicz, P. Discrete model of opinion changes using knowledge and emotions as control variables. *PloS One* . (2012). 7: e44489. doi: 10. 1371/journal. pone. 0044489

96. Sobkowicz, P. Minority persistence in agent based model using information and emotional arousal as control variables. *Eur Phys J B* . (2013). 86: 1–11. doi: 10. 1140/epjb/e2013-40029-0

97. Sobkowicz, P. Quantitative agent based model of opinion dynamics: polish elections of 2015. *PloS One* . (2016). 11: e0155098. doi: 10. 1371/journal. pone. 0155098.

98. Garcia, D, and Schweitzer, F. Emotions in product reviews–empirics and models. 2011 IEEE international conference on privacy, security, risk, and trust, and IEEE international conference on social computing ; 2011 Oct 9–11 ; Boston, MA . Piscataway, NJ: IEEE (2010). p. 483–8.

99. Tanase, D, Garcia, D, Garas, A, and Schweitzer, F. Emotions and activity profiles of influential users in product reviews communities. *Front Phys* . (2015). 3: 87. doi: 10. 3389/fphy. 2015. 00087.

100. Garcia, D, Kappas, A, Küster, D, and Schweitzer, F. The dynamics of emotions in online interaction. *R Soc Open Sci* . (2016). 3: 160059. doi: 10. 1098/rsos. 160059.

101. Schweitzer, F, Krivachy, T, and Garcia, D. An agent-based model of opinion polarization driven by emotions. *Complexity* . (2020). 2020: 5282035. doi: 10. 1155/2020/5282035.

102. Tversky, A, and Kahneman, D. Probabilistic reasoning. In: Goldman, A, editor *Readings in philosophy and cognitive Science* . Cambridge, UK: The MIT Press (1993). p. 43–68.

103. Kahneman, D. *Thinking, fast and slow* . New York, NY: Macmillan (2011). 512 p.

104. Hilbert, M. Toward a synthesis of cognitive biases: how noisy information processing can bias human decision making. *Psychol Bull* . (2012). 138: 211–37. doi: 10. 1037/a0025940.

105. Benson, B. Cognitive bias cheat sheet (2016). Available from: https://betterhumans. coach. me/cognitive-bias-cheat-sheet-55a472476b18 (accessed 07 24, 2020).

106. Sobkowicz, P. Opinion dynamics model based on cognitive biases. *J Artif Soc Soc Simulat* . (2018). 21: 8. doi: 10. 18564/jasss. 3867

107. Geschke, D, Lorenz, J, and Holtz, P. The triple-filter bubble: using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *Br J Soc Psychol* . (2019). 58: 129–49. doi: 10. 1111/bjso. 12286

108. Epstein, JM. Why model? *J Artif Soc Soc Simulat* . (2008). 11: 12.

109. Thompson, NS, and Derr, P. Contra Epstein, good explanations predict. *J Artif Soc Soc Simulat* . (2009). 12: 9.

110. Troitzsch, KG. Not all explanations predict satisfactorily, and not all good predictions explain. *J Artif Soc Soc Simulat* . (2009). 12: 10.

111. Nowak, A, Rychwalska, A, and Borkowski, W. Why simulate? to develop a mental model. *J Artif Soc Soc Simulat* . (2013). 16: 2235. doi: 10. 18564/jasss. 2235.

112. Kasim, M, Watson-Parris, D, Deaconu, L, Oliver, S, Hatfield, P, Froula, D, et al. Up to two billion times acceleration of scientific simulations with deep

neural architecture search. Preprint repository name [Preprint] (2020). Available from: https://arxiv. org/abs/2001. 08055

113. Hofman, JM, Sharma, A, and Watts, DJ. Prediction and explanation in social systems. *Science* . (2017). 355: 486–8. doi: 10. 1126/science. aal3856.

114. Jungherr, A, and Theocharis, Y. The empiricist's challenge: asking meaningful questions in political science in the age of big data. *J Inf Technol Polit* . (2017). 14: 97–109. doi: 10. 1080/19331681. 2017. 1312187.

115. Zehmakan, AN, and Galam, S. Rumor spreading: a trigger for proliferation or fading away. *Chaos* . (2020). 30: 073122. doi: 10. 1063/5. 0006984

116. Boschi, G, Cammarota, C, and Kühn, R. Opinion dynamics with emergent collective memory: a society shaped by its own past. *Phys Stat Mech Appl* . (2020). 558: 124909. doi: 10. 1016/j. physa. 2020. 124909.

117. An, X, Ding, L, and Hu, P. Information propagation with individual attention-decay effect on activity-driven networks. *Phys Stat Mech Appl* . (2020). 556: 12815. doi: 10. 1016/j. physa. 2020. 124815.

118. Bond, RM, Fariss, CJ, Jones, JJ, Kramer, ADI, Marlow, C, Settle, JE, et al. A 61-million-person experiment in social influence and political mobilization. *Nature* . (2012). 489: 295–8. doi: 10. 1038/nature11421.

119. Sobkowicz, P. Social simulation models at the ethical crossroads. *Sci Eng Ethics* . (2019). 25: 143–57. doi: 10. 1007/s11948-017-9993-0

120. Schweitzer, F, Krivachy, T, and Garcia, D. How emotions drive opinion polarization: an agent-based model. Preprint repository name [Preprint] (2019). Available from: https://arxiv. org/abs/1908. 11623

121. Das, R, Kamruzzaman, J, and Karmakar, G. Modelling majority and expert influences on opinion formation in online social networks. *World Wide Web* . (2018). 21: 663–85. doi: 10. 1007/s11280-017-0484-7.

122. Kahan, DM, Jenkins-Smith, H, and Braman, D. Cultural cognition of scientific consensus. *J Risk Res* . (2011). 14: 147–74. doi: 10. 1080/13669877. 2010. 511246.

123. Barrera Lemarchand, F, Semeshenko, V, Navajas, J, and Balenzuela, P. Polarizing crowds: consensus and bipolarization in a persuasive arguments model. *Chaos* . (2020). 30: 063141. 10. 1063/5. 0004504

124. Twenge, JM. *iGen. Why today's super-connected kids are growing up less rebellious, more tolerant, less happy–and completely unprepared for adulthood–and what that means for the rest of us* . New York, NY: Simon & Schuster (2017). 352 p.

125. De Nooy, W, and Kleinnijenhuis, J. Polarization in the media during an election campaign: a dynamic network model predicting support and attack among political actors. *Polit Commun* . (2013). 30: 117–38. doi: 10. 1080/10584609. 2012. 737417.

126. Ito, J, and Kaneko, K. Spontaneous structure formation in a network of dynamic elements. *Phys Rev* . (2003). 67: 046226. doi: 10. 1103/physreve. 67. 046226.

127. Liu, C, and Zhang, Z-K. Information spreading on dynamic social networks. *Commun Nonlinear Sci Numer Simulat* . (2014). 19: 896–904. doi: 10. 1016/j. cnsns. 2013. 08. 028.

128. Perra, N, Gonçalves, B, Pastor-Satorras, R, and Vespignani, A. Activity driven modeling of time varying networks. *Sci Rep* . (2012). 2: 469. doi: 10. 1038/srep00469.

129. Twain, M. Running for governor. The Galaxy (1870). Available at: http://twainquotes. com/Galaxy/187012c. html.

130. Bessi, A, Zollo, F, Del Vicario, M, Scala, A, Caldarelli, G, and Quattrociocchi, W. Trend of narratives in the age of misinformation. *PloS One* . (2015). 10: e0134641. doi: 10. 1371/journal. pone. 0138740.

131. Bessi, A, Petroni, F, Del Vicario, M, Zollo, F, Anagnostopoulos, A, Scala, A, et al. *Proceedings of the 24th international conference on world wide Web companion* . International World Wide Web Conferences Steering Committee (2015b). p. 355–6. Viral misinformation: the role of homophily and polarization

132. Zollo, F, and Quattrociocchi, W. Misinformation spreading on facebook. In: *Complex spreading phenomena in social systems* . Springer (2018). p. 177–96.

133. Del Vicario, M, Bessi, A, Zollo, F, Petroni, F, Scala, A, Caldarelli, G, et al. Echo chambers in the age of misinformation. Preprint repository name [Preprint] (2015). Available from: https://arxiv. org/abs/1509. 00189

134. Del Vicario, M, Bessi, A, Zollo, F, Petroni, F, Scala, A, Caldarelli, G, et al. . The spreading of misinformation online. *Proc Natl Acad Sci* . (2016). 113: 554–9. doi: 10. 1073/pnas. 1517441113

135. Berinsky, AJ. Rumors and health care reform: experiments in political misinformation. *Br J Polit Sci* . (2017). 47: 241–62. doi: 10. 1017/s0007123415000186.

136. Swire, B, Berinsky, AJ, Lewandowsky, S, and Ecker, UK. Processing political misinformation: comprehending the trump phenomenon. *R Soc Open Sci* . (2017). 4: 160802. doi: 10. 1098/rsos. 160802.

137. Zollo, F, Novak, PK, Del Vicario, M, Bessi, A, Mozetič, I, Scala, A, et al. . Emotional dynamics in the age of misinformation. *PloS One* . (2015). 10: e0138740. doi: 10. 1371/journal. pone. 0138740.

138. Zollo, F, Bessi, A, Del Vicario, M, Scala, A, Caldarelli, G, Shekhtman, L, et al. . Debunking in a world of tribes. Preprint repository name [Preprint] (2015). Available from: https://arxiv. org/abs/1510. 04267

139. Hardaker, C. Trolling in asynchronous computer-mediated communication: from user discussions to academic definitions. *J Politeness Res* . (2010). 6: 215–42. doi: 10. 1515/jplr. 2010. 011.

140. Birkbak, A. Into the wild online: learning from Internet trolls. *Clin Hemorheol and Microcirc* . (2018). 23: 8297. doi: 10. 5210/fm. v23i5. 8297

141. Zelenkauskaite, A, and Niezgoda, B. " Stop Kremlin trolls:" Ideological trolling as calling out, rebuttal, and reactions on online news portal

commenting. *Clin Hemorheol and Microcirc* . (2017). 22: 7795. doi: 10. 5210/fm. v22i15. 7795

142. Bradshaw, S, and Howard, P. Troops, trolls and troublemakers: a global inventory of organized social media manipulation. Tech. Rep. Oxford, UK: Oxford Internet Institute (2017).

143. Kurowska, X, and Reshetnikov, A. Neutrollization: industrialized trolling as a pro-kremlin strategy of desecuritization. *Secur Dialog* . (2018). 49: 345–63. doi: 10. 1016/j. paid. 2014. 01. 016.

144. Buckels, EE, Trapnell, PD, and Paulhus, DL. Trolls just want to have fun. *Pers Indiv Differ* . (2014). 67: 97–102.

145. Mihaylov, T, Koychev, I, Georgiev, G, and Nakov, P. Exposing paid opinion manipulation trolls. In: Proceedings of the international conference recent advances in natural language processing ; 2015 Sep ; Hissar, Bulgaria . Shoumen, Bulgaria: INCOMA Ltd (2015). p. 443–50.

146. De La Vega, L, and Ng, V. Determining trolling in textual comments. In: 11th international conference on language resources and evaluation ; Miyazaki, Japan . Miyazaki, Japan: Phoenix Seagaia Conference Center (2018). p. 3701–6.

147. Himelboim, I, Gleave, E, and Smith, M. Discussion catalysts in online political discussions: content importers and conversation starters. *J Computer-Mediated Commun* . (2009). 14: 771–89. doi: 10. 1111/j. 1083-6101. 2009. 01470. x.

148. Kelly, J, Fisher, D, and Smith, M. Friends, foes, and fringe: norms and structure in political discussion networks. In: Davies, T, and Gangadharan, SP, editors *Online deliberation: design, research, and practice* . CSLI Publications/University of Chicago Press (2009). p. 83–93.

149. Martins, ACR. Continuous opinions and discrete actions in opinion dynamics problems. *Int J Mod Phys C* . (2008). 19: 617–24. doi: 10. 1142/s0129183108012339.

150. Martins, AC. Discrete opinion models as a limit case of the CODA model. *Physica A* . (2014). 395: 352–7. doi: 10. 1016/j. physa. 2013. 10. 009

151. Jedrzejewski, A, Marcjasz, G, Nail, PR, and Sznajd-Weron, K. Think then act or act then think? *PloS One* . (2018). 13: e0206166. doi: 10. 1371/journal. pone. 0206166.

152. Furnham, A, Eracleous, A, and Chamorro-Premuzic, T. Personality, motivation and job satisfaction: hertzberg meets the big five. *J Manag Psychol* . (2009). 24: 765–79. doi: 10. 1108/02683940910996789.

153. Haidt, J, and Joseph, C. The moral mind: how five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In: Carruthers, P, Laurence, S, and Stich, S, editors *The innate mind* . New York, NY: Oxford University Press (2007). p. 367–91.

154. Haidt, J. The new synthesis in moral psychology. *Science* . (2007). 316: 998–1002. doi: 10. 1126/science. 1137651.

155. Graham, J, Haidt, J, Koleva, S, Motyl, M, Iyer, R, Wojcik, S, et al. . Moral foundations theory: the pragmatic validity of moral pluralism. *Adv Exp Soc Psychol* . (2013). 47: 55–130. doi: 10. 1016/b978-0-12-407236-7. 00002-4.

156. Bruner, J. The narrative construction of reality. *Crit Inq* . (1991). 18: 1–21. doi: 10. 1086/448619.

157. White, H. The value of narrativity in the representation of reality. *Crit Inq* . (1980). 7: 5–27. doi: 10. 1086/448086.

158. Graesser, AC, Olde, B, and Klettke, B. Narrative impact: Social and cognitive foundations. In: Green,, MC, and Strange,, JJBrock, TC, editors *How does the mind construct and represent stories* . Oxfordshire, UK: Taylor & Francis (2002). p. 229–62.

159. Bessi, A, Coletto, M, Davidescu, GA, Scala, A, Caldarelli, G, and Quattrociocchi, W. Science vs conspiracy: collective narratives in the age of misinformation. *PloS One* . (2015). 10: 02. doi: 10. 1371/journal. pone. 0118093.

160. Tversky, A, and Kahneman, D. Judgment under uncertainty: heuristics and biases. *Science* . (1974). 185: 1124–31. doi: 10. 1126/science. 185. 4157. 1124.

161. Wojcieszak, M, and Kim, N. How to improve attitudes toward disliked groups: the effects of narrative versus numerical evidence on political persuasion. *Commun Res* . (2016). 43: 785–809. doi: 10. 1177/0093650215618480.

162. Wojcieszak, M, Azrout, R, Boomgaarden, H, Alencar, AP, and Sheets, P. Integrating Muslim immigrant minorities: the effects of narrative and statistical messages. *Commun Res* . (2017). 44: 582–607. doi: 10. 1177/0093650215600490.

163. Kunda, Z. The case for motivated reasoning. *Psychol Bull* . (1990). 108: 480. doi: 10. 1037/0033-2909. 108. 3. 480.

164. Redlawsk, DP. Hot cognition or cool consideration? testing the effects of motivated reasoning on political decision making. *J Polit* . (2002). 64: 1021– 44. doi: 10. 1111/1468-2508. 00161.

165. Bolsen, T, Druckman, JN, and Cook, FL. The influence of partisan motivated reasoning on public opinion. *Polit Behav* . (2014). 36: 235–62. doi: 10. 1007/s11109-013-9238-0.

166. Yeo, PDSK, Cacciatore, MA, and Scheufele, DA. News selectivity and beyond: motivated reasoning in a changing media environment. In: Jandura, O, Petersen, T, Mothes, C, and Schielicke, AM, editors *Publizistik und gesellschaftliche Verantwortung* . Wiesbaden, Germany: Springer (2015). p. 83–104.

167. Kahan, DM. The politically motivated reasoning paradigm, part 1: what politically motivated reasoning is and how to measure it. In: Scott, R, and Kosslyn, S, editors *Emerging trends in the social and behavioral sciences* . New York, NY: Wiley Online Library (2016). p. 1–16.

168. Kahan, DM. The politically motivated reasoning paradigm, part 2: unanswered questions. In: Scott, R, and Kosslyn, S, editors *Emerging trends*

in the social and behavioral sciences . New York, NY: Wiley Online Library (2016). p. 1–15.

169. Baekgaard, M, Christensen, J, Dahlmann, CM, Mathiasen, A, and Petersen, NBG. The role of evidence in politics: motivated reasoning and persuasion among politicians. *Br J Polit Sci* . (2019). 49: 1117–40. doi: 10. 1017/s0007123417000084.

170. Pennycook, G, and Rand, DG. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *J. Personality* . (2019). 88. doi: 10. 1111/jopy. 12476.

171. McKnight, DH, and Chervany, NL. Trust and distrust definitions: one bite at a time. In: Falcone, R, Singh, M, and Tan, YH, editors *Trust in cyber-societies* . *Lecture Notes in Computer Science* , Vol. 2246. Berlin, Heidelberg: Springer (2001). p. 27–54.

172. Renn, O, and Levine, D. Credibility and trust in risk communication. In: Kasperson, RE, and Stallen, PJM, editors. *Communicating risks to the public. Technology, risk, and society* . Dordrecht, Netherlands: Springer (1991). p. 175–217.

173. Slovic, P. Trust, emotion, sex, politics, and science: surveying the risk-assessment battlefield. *Risk Analysis* . (1999). 19: 689–701. doi: 10. 1111/j. 1539-6924. 1999. tb00439. x.

174. Krawczyk, M, Malarz, K, Korff, R, and Kulakowski, K. Communication and trust in the bounded confidence model. In: Pan, JS, Chen, SM, and Nguyen, NT, editors *Computational collective intelligence. technologies and*

applications. ICCCI 2010 . *Lecture Notes in Computer Science* , Vol. 6421. Berlin, Heidelberg: Springer (2010). p. 90–9.

175. Pan, Z. Trust, influence, and convergence of behavior in social networks. *Math Soc Sci* . (2010). 60: 69–78. doi: 10. 1016/j. mathsocsci. 2010. 03. 007

176. Quandt, T. What's left of trust in a network society? an evolutionary model and critical discussion of trust and societal communication. *Eur J Commun* . (2012). 27: 7–21. doi: 10. 1177/0267323111434452.

177. Tsang, A, and Larson, K. Opinion dynamics of skeptical agents. In: 13th international conference on autonomous agents and multiagent systems (AAMAS 2014) ; 2014 May ; Paris, France . AAMAS (2014). p. 277–84.

178. Yi, C, Bao, Y, Jiang, J, and Xue, Y. Modeling cascading failures with the crisis of trust in social networks. *Phys Stat Mech Appl* . (2015). 436: 256–71. doi: 10. 1016/j. physa. 2015. 05. 030.

179. Fan, K, and Pedrycz, W. Emergence and spread of extremist opinions. *Phys Stat Mech Appl* . (2015). 436: 87–97. doi: 10. 1016/j. physa. 2015. 05. 056.

180. Xiong, F, Liu, Y, and Cheng, J. Modeling and predicting opinion formation with trust propagation in online social networks. *Commun Nonlinear Sci Numer Simulat* . (2017). 44: 513–24. doi: 10. 1016/j. cnsns. 2016. 09. 015.

181. Nowak, A, Vallacher, R, Rychwalska, A, Roszczyńska-Kurasińska, M, Ziembowicz, K, Biesaga, M, et al. . *Target in control social influence as distributed information processing* . Berlin, Germany: Springer (2020). 89 p.

182. Ishii, A, and Kawahata, Y. Theory of opinion distribution in human relations where trust and distrust mixed. In *Intelligent decision technologies* . Berlin, Germany: Springer (2020). p. 471–8.

183. Mulder, LB, Van Dijk, E, De Cremer, D, and Wilke, HA. Undermining trust and cooperation: the paradox of sanctioning systems in social dilemmas. *J Exp Soc Psychol* . (2006). 42: 147–62. doi: 10. 1016/j. jesp. 2005. 03. 002.

184. Balliet, D, and Van Lange, PAM. Trust, conflict, and cooperation: a meta-analysis. *Psychol Bull* . (2013). 139: 1090–112. doi: 10. 1037/a0030939

185. Evans, AM, and Krueger, JI. The psychology (and economics) of trust. *Soc Person Psychol Compass* . (2009). 3: 1003–17. doi: 10. 1111/j. 1751-9004. 2009. 00232. x.

186. Bierhoff, H-W, and Vornefeld, B. The social psychology of trust with applications in the Internet. *Analyse und Kritik* . (2004). 26: 48–62. doi: 10. 1515/auk-2004-0103.

187. Rapp, C. Moral opinion polarization and the erosion of trust. *Soc Sci Res* . (2016). 58: 34–45. doi: 10. 1016/j. ssresearch. 2016. 02. 008.

188. LaCour, MJ, and Green, DP. Retracted article: when contact changes minds: an experiment on transmission of support for gay equality. *Science* . (2014). 346: 1366–9. doi: 10. 1126/science. 1256151.

189. Broockman, D, Stanford, G, and Aronow, P. Irregularities in LaCour (2014) (2015). Available at: https://osf. io/preprints/metaarxiv/qy2se/ .

190. Broockman, D, and Kalla, J. Durably reducing transphobia: a field experiment on door-to-door canvassing. *Science* . (2016). 352: 220–4. doi: 10. 1126/science. aad9713.

191. Panger, G. Reassessing the facebook experiment: critical thinking about the validity of big data research. *Inf Commun Soc* . (2016). 19: 1108–26. doi: 10. 1080/1369118x. 2015. 1093525.