Article

Opinion Polarization in Human Communities Can Emerge as a Natural Consequence of Beliefs Being Interrelated

Anna Zafeiris [1,2]

1 MTA-ELTE Statistical and Biological Physics Research Group, Pázmány Péter Stny. 1/A,
1117 Budapest, Hungary; anna.kinga.zafeiris@ttk.elte.hu
2 MTA-ELTE 'Lendület' Collective Behaviour Research Group, Hungarian Academy of Sciences,
Eötvös University, 1117 Budapest, Hungary

Abstract: The emergence of opinion polarization within human communities—the phenomenon that individuals within a society tend to develop conflicting attitudes related to the greatest diversity of topics—has been a focus of interest for decades, both from theoretical and modelling points of view. Regarding modelling attempts, an entire scientific field—opinion dynamics—has emerged in order to study this and related phenomena. Within this framework, agents' opinions are usually represented by a scalar value which undergoes modification due to interaction with other agents. Under certain conditions, these models are able to reproduce polarization—a state increasingly familiar to our everyday experience. In the present paper, an alternative explanation is suggested along with its corresponding model. More specifically, we demonstrate that by incorporating the following two well-known human characteristics into the representation of agents: (1) in the human brain beliefs are interconnected, and (2) people strive to maintain a coherent belief system; polarization imme- diately occurs under exposure to news and information. Furthermore, the model accounts for the proliferation of fake news, and shows how opinion polarization is related to various cognitive biases.

Keywords: belief systems; opinion polarization; belief system dynamics; fake news; cognitive biases; opinion dynamics

1. Introduction

As Evan Williams, co-founder of Twitter, famously said in a 2017 interview, "I thought once everybody could speak freely and exchange information and ideas, the world is automatically going to be a better place. I was wrong about that" [1]. This is a good overview of the surprise that the unprecedented connectivity among people—primarily driven by various internet-based social media platforms in the early 21st century—brought us unprecedented factions, dissension and fake news [2,3], instead of agreement and conciliation [4,5]. The reasons behind these phenomena are diverse and manifold and, accordingly, are the subject of the most diverse scientific fields from history [6] through sociology [7] to computational social science [8–10].

The quest for finding a valid explanation—and a practicable model—for the phe- nomena of the above mentioned polarization and fragmentation has been underway for decades. Specifically, regarding computational models, an entire field, *opinion dynamics* , has emerged in order to study the way opinions, information, views and beliefs propagate in human communities [11–16]. In general, these models study the dynamics of attitudes related to a certain topic, such as issues related to climate change, abortion, immigration, vaccination, a certain politician or political party, etc. Typically, the attitudes of the agents towards the given issue are described by scalar values which are assumed to be altered due to communication with peers [11,12,17]. Within this scientific field, *consensus* refers to the state in which all social actors share the same opinion, *polarization* to the condition when

each agent accepts one of two opposing opinions, while *fragmentation* refers to a transitory phase, in which a finite number of distinct "opinion islands" appear.

A key concept related to continuous opinion dynamics models—models representing opinions as scalars taken from a continuous interval usually between -1 and $+1$—is the "confidence threshold", which is a value above which agents cease to communicate with each other. Above this threshold, "classical" bounded confidence models predict consensus, and fragmentation and/or polarization under it [13–15,17]. However, the polarization/fragmentation breaks down in the presence of noise [18–22]—an essential and unavoidable component in all social and biological systems [17]. In order to reach polarization, other mechanisms have been suggested, such as "distancing", which is the direct amplification of differences between dissimilar individuals [18]. Lately, these models have been further developed by incorporating more realistic features, such as heterogeneity with respect to the agents' confidence thresholds [23] or by incorporating an underlying communication network [24] defining the pattern by which social actors interact with each other. These approaches lead to more complex—and realistic—dynamics. Furthermore, recently, building on the observation that *events* often have a polarization effect [25,26], the process of polarization was modelled by extending a classical bounded confidence model—the so called Hegselmann-Krause model [13]—in a way that individuals change their opinions in line with the certain event [27]. Other models approach the problem from a kinetic [28] or hydrodynamic [29] perspective.

In the present paper, an alternative, or complementary, explanation is suggested, by showing that in case some "basic" human characteristics are incorporated, polarization immediately appears once agents are exposed to new information—even without direct communication. These almost trivial human characteristics are that (i) human beliefs are interrelated rather then evolving independently of each other, and (ii) people strive to maintain a coherent, contradiction-free belief-system. The fundamental difference between "classical" opinion dynamics models and the ones incorporating such features is that while in the first case agents are represented with a single scalar value in the second case the model of the social actors has some kind of inner structure. This difference gives rise to an entirely different dynamic [30,31].

2. The Background: Fundamental Features of Human Belief Systems

During the last decades, a vast amount of knowledge has accumulated in scientific fields on the ways humans perceive the world, make decisions and structure their beliefs. Despite the fact that obtaining a detailed understanding of these processes still require further work, some scientific fields, such as neurobiology [32–36] or various human sciences [37,38], such as psychology [39–42], anthropology [43], economics [44–46] and political science [4,5,47–51] have progressed considerably. From our point of view, the key finding is that in humans, opinions and beliefs *never* occur alone, that is, no concept or belief can exist in isolation. (Actually, humans are not even able to memorize anything without connecting it to something meaningful [32].) Rather, concepts and beliefs are organized into a structure, a system (*"belief system"*) which has well-defined features [32,47,48,52]:

1. *First* and fore-most, it seek to be *consistent*. This means that people try to maintain a belief system in which the elements mutually support each other, or are independent [48,52]. In the case of holding *conflicting* beliefs, people experience discomfort called *cognitive dissonance* [53,54] which people will try to reduce. By this time, cognitive dissonance has become one of the most influential and well-researched theories in social psychology [54–56].

2. *Secondly*, beliefs are not equally important: those that are more personal, closer to the "self", "identity" or "ego", trigger more intense feelings and are more difficult to change [43,46,54,57]. In this sense, beliefs have a *hierarchical property*: the ones that are higher in rank define or constrain the ones that are lower in rank [58]. For example, the belief or disbelief in God is a central (high-ranking) element in one's belief system, while the belief that "an egg should be boiled for seven minutes in order to get the

best soft-boiled egg" is a low ranking one, and, accordingly, can be changed more easily [43]. Furthermore, beliefs that people hold in high regard tend to cause greater dissonance in case of contradiction with other beliefs [54].

3. *Finally*, beliefs belonging to the same broader topic (e.g., health, art-related topics, religion, political issues, etc.) are more strongly interrelated than beliefs belonging to different topics. For example, attitudes towards "freedom of speech", "religious freedom" and "freedom to choose spouse" are more closely related than beliefs regarding "freedom of speech" and, say, homeopathic treatments. In mathematical (graph-theoretical) terms, belief systems are "modular" or "compartmentalized" [43,50].

As a first approximation, such a structure can be represented as a *modular, hierarchical network* (graph) in which the nodes are connected by supportive (positive) or invalidating (negative) relations [48–50,59] (Figure 1). For example, according to a widespread belief, "Pathogens can cause diseases". According to another wide-spread belief—primarily in historical societies—"Diseases are caused by evil spirits" [60,61]. These two concepts are in negative relation: somebody believing in one of these opinions will probably disagree with the other. In contrast, the beliefs "Pathogens can cause diseases" and "Contagion is due to the spread of pathogens" support each other (positive relationship), since accepting one of them renders the acceptance of the other more probable.

This last requirement ensures that the *level of consistency* can be defined [62]. In this description, nodes are beliefs, and edges represent functional relationships between them [52,59,63]. This approach has already been applied by sociologists and economists as well in order to model political belief system dynamics [49,50,59,63,64]. Within this framework, the focus is on the *relatedness* of beliefs—captured by graph representation—while the hierarchical and modular characteristics do not gain special importance (see the Supplementary Material for the results incorporating the hierarchical characteristics as well).
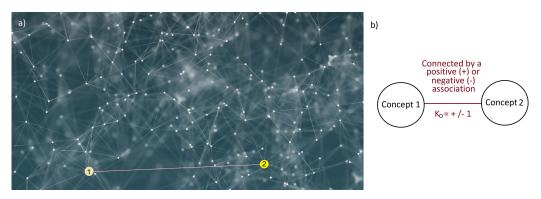

Figure 1. The graph representation of belief systems. (**a**) As a first approximation, human belief systems can be represented by networks in which nodes are beliefs ("elements of a belief system" [59]) and edges represent relationships. (**b**) Links (relationships) can be either supportive (positive) or contradictory (negative) [62].

In these models, a node is "an element of a person's belief system" [59], which the related literature names slightly differently, such as "opinions" [49], "concepts" [62] "attitudes" [49,59,64], "beliefs" [50,59,62], or "positions" [63,64]. In the present article, the terms *"belief"*, *"attitude"* and *"concept"* are mostly used, under the condition that we consider all sorts of human thoughts as an "element of a belief system" (that is: a "belief"), whether they be simple or complex, that can be transmitted with the use of language from one person's mind to another's [38,43]. Accordingly, a node can be a thought or attitude towards public issues such as abortion, climate change, immigration, vaccination, gun control; it can be information regarding a public figure or a political party, an idea related to "proper behaviour", "justice"; or the belief that somebody did or said something, etc. Furthermore, this definition implies that beliefs can be transmitted via communication. Communication,

in its most basic form, can be discussion (talking) between two or more individuals, but it can also be news/beliefs/information spread by a single agent or organization to many people at the same time, for example via public and social media, news channels, journals, etc. In short, communication is the circulation of news, information and beliefs within a certain community.

The fact that beliefs (attitudes/concepts) are in functional relation with each other (that is, if they are connected, they either support or contradict each other) is crucial, because it implies that some "new" belief will fit into an already existing system—the ones that *in*crease the system's consistency—while others—the ones *de*creasing the system's consistency—will not [52]. These latter gives rise to the disturbing feeling of cognitive dissonance and people will apply various strategies in order to avoid them. Among other strategies, they will try to keep contact only with those from whom they expect reassuring information (homophily), they will ignore certain information and focus on other information (attentional bias), while greater credence will be given to evidence that fits with the existing beliefs (confirmation bias). These strategies are known as various *biases* in the field of psychology and sociology [40,65–67]. Different people apply different strategies to various extents; however, to some level, all the strategies are applied by all of us [40,41,65,66].

Furthermore, these cognitive dissonance avoiding mechanisms are in close relation with the proliferation of fake news and the circulation of various types of questionable information as well. In case the well-fitting of a piece of information into the already existing belief system weights more than its credibility, people will adopt it—simply because it provides the pleasant feeling of reassurance. This mechanism is applied in all aspects of life, not only in case of political issues. For example, in the field of economics, it has been observed that managers, whose sales fall short of expectations, rather than rethinking the qualities of the product, tend to identify the cause of the failure elsewhere, for example in the marketing campaign. In such cases, they state that the *marketing* campaign failed, so it is actually a miracle that the product was sold at all [68]. By finding this new explanation, the sales results show directly the *merits* of the product, not its failure, and as such, serves as a basis for the pleasant feeling of reassurance. This mechanism is analyzed in Section 4.2.

3. The Model

We assume a population of N agents. At this point, only their attitudes towards two concepts are important (the attitudes towards, say, vaccination (Concept 1) and, say, a certain public figure (Concept 2)). People can hold any kind of attitudes towards these concepts, from total condemnation (marked by -1) to total support (denoted by $+1$). Neutrality or indifference is indicated by zero or near-zero values. We are interested in how the agents' attitudes evolve due to being exposed to some news (piece of information) that creates a relation between two originally independent concepts (see Figure 1b).

The relation K_0 can be positive or negative. Using the above example, a trivial positive connection can be that "XY public figure (concept 2) has spoken out *in favor* of vaccination (concept 1)" ($K_0 = +1$), while a negative connection can be that "XY public figure has spoken out *against* it" ($K_0 = -1$).

In case an agent holds positive attitudes towards both concepts, the positive message (support of vaccination) will give rise to the comforting feeling of *reassurance* [41]. In this case, the original attitudes are reinforced, since both concepts become better connected and further embedded into the belief system. In contrast, in case the XY trusted and respected politician takes a position *against* vaccination, a supported matter, the agent will experience *cognitive dissonance* with an intensity proportional to the original attitude values, and will apply a strategy in order to reduce it [53,54].

Turning back to the basic scenarios, an agent can hold negative attitude towards one of the concepts, and a positive one towards the other—say, for example, a negative attitude towards the public figure and positive attitude towards vaccination. In this case, a negative relation will give rise to reassurance, i.e., "XY politician, whom I anyway hold very low,

talked out against vaccination, a cause so important for me ... No surprise here, a fool is known by his conversation" and so on. All scenarios can be analyzed with the same train of thought.

Accordingly, from a mathematical point of view, the cognitive dissonance (or reassurance), $C_i(t)$, that agent i will experience at time-step t, can be formulated as:

$$C_i(t) = a_{i,1}(t) \cdot a_{i,2}(t) \cdot K_0 \tag{1}$$

where $a_{i,1}(t)$ and $a_{i,2}(t)$ are the original attitudes of agent i towards concept 1 and 2, respectively, at time-step t, and K_0 is the type of connection, which can take two values, $+1$ or -1, according to the supportive or opposing nature of the connection between the concepts (see also Figure 1b). In case C is positive, it is called *reassurance*, while in case it is negative, it is usually referred to as *cognitive dissonance*. Anyhow, in both cases, C denotes the value by which the information alters the *coherence* or *consistency* level of agent i's belief system. Note that $| C_i(t) | \leq 1$ is always the case, since $| a_{i,1}(t) | \leq 1$, $| a_{i,2}(t) | \leq 1$, and $K_0 = \pm 1$.

According to the literature [40,41,65,66], in case of facing information inducing cognitive dissonance, people attempt to relieve the discomfort in different ways, among which the most common ones are:

(i)   Rejecting new information that conflicts with the already existing ones;
(ii)   Re-evaluating the attitudes;
(iii) A tendency of "explaining things away", that is, finding alternative explanations (developing new beliefs) which supplement the original information in a way that the primordial contradiction is dissolved.

From a modeling point of view, the first strategy—rejecting the information—simply leaves the belief system unaltered. In this case, in the framework of the model, the network—nodes, edges and weights—remain unchanged. The second and third strategies do modify the belief system, due to the new connection between the originally unconnected concepts. In the following section, we will focus on modelling these strategies.

3.1. Modelling the Re-Evaluation of Beliefs

The constant re-evaluation of our already existing beliefs is an inevitable part of the process of learning and development [32]. New information often comes in the form of creating connection among concepts and beliefs that were originally disconnected. As a matter of fact, this is a basic form of learning. Furthermore, people tend to evaluate most information, beliefs and concepts according to some personal narrative, a personal "frame of mind", which is different from person to person. Simply put, this unique narrative is our personality [37], which defines the very way we perceive the world and make decisions [45,46]. This variety entails individual differences in evaluating the most diverse topics around us, whether it be the judgement of a public figure, a movie or the question of immigration.

In the context of a formal model, the most simple and plausible way to grasp these attitudes is to use numbers between -1 and $+1$ in a way that negative values represent negative attitudes and positive ones refer to positive stances. The two extreme values, -1 and $+1$, refer to complete condemnation/approval, respectively.

In order to see *how* these values might change, consider for example the following case: Paul believes that, say, genetically modified food is harmful. He has already heard it from his friends, and now he reads it in his favorite blog as well. This gives him a feeling of reassurance, due to which he will be convinced about the verity of this belief even more, and will be more attached to his favorite blog as well. In other words, the "embeddedness" of the original attitudes will increase. Mathematically speaking, his already positive attitudes towards these concepts (his belief and the blog) will increase even more due to the positive connection. Now consider a situation where he learns the opposite from his favorite blog, namely that there is nothing at all that could be harmful in genetically

modified food (that is, a *negative* association appears among the two positive concepts: the belief and the blog). In this case, he will experience some level of cognitive dissonance, whose extent depends on his original commitments towards the two concepts [42,53,56]. This experience will make him less convinced, either of the reliability of the blog or of the belief itself—or both. Mathematically speaking, the originally positive values (attached to the two concepts) will decrease somewhat. In other words, *cognitive dissonance* (negative $C_i(t)$ values) *de*creases the absolute value of the affected attitude (k), while *reassurance* (positive C values) *in*creases it. Consider the following formula:

$$a_{i,k}(t+1) = sign(a_{i,k}(t)) \cdot (\mid a_{i,k}(t) \mid + \rho \cdot C_i(t)) + Z_A \qquad (2)$$

where $a_{i,k}(t)$ is the original attitude of agent i at time-step t towards attitude k, $sign(a_{i,k}(t))$ is its signal ($+$ or -1), ρ is a random value ("noise") taken from the $[0,1]$ interval with uniform distribution, effecting the extent to which the attitude changes, and $C_i(t)$ (defined by Equation (1)), is the level of "coherence" (commonly known as *cognitive dissonance*, in case it is negative, and *reassurance* in case it is positive). Finally, the Z_A noise comprises the effects of other factors influencing the change of attitudes. It can be either positive or negative with equal probability. In case the updated attitude value $a_{i,k}(t+1)$ falls outside the predefined $[-1,1]$ interval, it is set to the nearest threshold ($+1$ or -1).

Attitudes do not vary with the same probability and to the same extent in case of different people and topics; for some, environmental issues are extremely important (and "nothing can change" this stance), some people are detached, while others are convinced that they are just evil-minded hoaxes. The more extreme an attitude is, the more difficult is to change it [43,46,54,57]. (See also Section 2, 2nd bulleted point, "hierarchical property" of belief systems).

Mathematically speaking, the feature *"more difficult to change"* can be introduced into the model in two ways:

1. The more extreme an attitude value a is, the lower the *probability* that it will change. Equation (3) expresses the most simple mathematical formulation of this relation.
2. The more extreme an attitude value a is, the smaller the *magnitude* with which it can change.

For the results presented in the main text, the above mentioned hierarchical property was introduced into the model according to the first way, that is, by setting the *probability* $p(AttCh_{i,k}(t))$ of attitude-change according to Equation (3), and setting ρ—the parameter controlling the maximal *extent* with which the attitude values alter due to the experienced cognitive dissonance or reassurance (C)—to 1. In other words, in Equation (2), $\rho = 1$ for the results presented in the main text. In the Supplementary Material, a detailed analysis is provided on how the parameter ρ effects the simulations (leading to the conclusions that the main claims remain valid, independently of the maximal extent of the alterations, see Figure S7).

$$p(AttCh_{i,k}(t)) = 1 - |a_{i,k}(t)| \qquad (3)$$

Note that in all the equations, the updated attitude values depend only on the agents' previous attitude values ($a_{i,k}(t)$), the type of the news (K_0), and on the cognitive dissonance (or reassurance) values that the news creates in the agents ($C_i(t)$). This means that agents develop their attitudes independently from each other. This originates from the fact that *the source of the information does not matter* in the present model. Accordingly, if the assumption is that it is the *agents* who circulate the news among themselves (for example in the form of "gossiping" either in person or on social media), then they interact with each other. In contrast, if the source of the information is something else (for example, state media or some kind of propaganda) then agents do not interact directly with each other. In reality, information usually circulates in both ways. The reason why entire populations are considered is twofold. Firstly, because one single agent cannot *"polarize"*; they can develop extreme attitudes under certain circumstances. *Polarization* is an emergent, statistical property of communities, a phenomenon which does not have an interpretation

on the level of individuals. The larger the statistics, the more apparent the phenomenon. The second reason is that after studying the elementary process of attitude-update in detail (which is the topic of the present paper), an immediate next step is to study the way by which agents manipulate and organize their social ties (links) assuming similar motivations (avoiding cognitive dissonance and enjoying reassurance).

3.2. Modelling the Inclusion of New Beliefs in Order to Relieve Cognitive Dissonance

In case a social actor experiences the upsetting feeling of cognitive dissonance due to a certain piece of information, a commonly applied strategy is to adopt—or create—an even newer belief that changes the context of the original one in a way that it does not serve as a basis of cognitive dissonance any longer; rather, it becomes neutral or even gives rise to the pleasant feeling of reassurance [40,41,68]. An example of this maneuver is mentioned at the end of Section 2, related to managers whose sales data lag behind the expectations tendentiously conceive of various explanations, e.g., ones related to "awfully managed" marketing campaigns. By adopting this new belief (namely that the marketing campaign was awfully managed), the cognitive dissonance caused by the negative sales results (linking a failure to their "self") is eliminated; furthermore, in this light, the sales-results could be seen as an achievement rather than a failure.

The most simple assumption is that the probability $p(NB_i(t))$ of adopting such a new belief (by agent i at time-step t) is proportional to the relief its adoption provides. Since only positive $C_i^{NB}(t)$ values represent reassurance, the most simple mathematical formula is the following:

$$p(NB_i(t)) = max(0, C_i^{NB}(t)) \tag{4}$$

where,

$$C_i^{NB}(t) = a_{i,1}(t) \cdot a_{i,3}(t) \cdot K_{NB} \tag{5}$$

As shown before, $a_{i,1}(t)$ and $a_{i,3}(t)$ are the attitudes of agent i towards concept 1 and the new belief at time-step t, respectively, and K_{NB} is the (positive or negative) connection type between them. Note that in case $C_i^{NB}(t)$ is negative—marking cognitive dissonance, instead of reassurance—the agent is highly unlikely to adopt the new belief.

4. Results

4.1. Re-Evaluating Beliefs

Let us consider a population in which the agents' initial attitudes towards two arbitrarily chosen concepts are distributed uniformly, taking values from the $[-1, 1]$ interval. In other words, at the beginning of the simulation, all sorts of attitudes are present in the population with equal probability, from complete condemnation to complete support and everything in between, with an average of zero. Let us now assume that this population is exposed to some kind of news, connecting the two originally unconnected concepts.

Assuming the most general setup, at each time-step t, a randomly chosen agent i acquires the information, and updates his/her attitudes according to Equation (2). (For the flowchart of the algorithm, see Figure A1a. The source of information can be anything, such as public or social media, propaganda, government information, etc. As it can be seen in Figure 2a,b,d,e, proportionally to the level of exposure (iteration number t), the attitudes tend to move towards the two extreme values, $+1$ and -1, either due to the experienced reassurance or due to the attempt to reduce cognitive dissonance (Equation (2)). The *distribution* of the attitude values within the population evolves very similarly in case of the two attitudes, since both are governed by Equation (2). (See Figure 2a,b,d,e). At high iteration numbers (indicating strong exposure to the news), around half of the population fully *supports* Concept 1—marked by attitude values close to $+1$—while the other crowd—composed of those whose attitude values are close to -1—fully *rejects* it. In other words, the population is *polarized* with respect to Concept 1 (Figure 2a,d). The same applies to Concept 2 (Figure 2b,e). In case the type of connection (K_0) is negative (Figure 2 bottom row), the two "stable points"—adopted by the vast majority of the population—are $(+1, -1)$

and $(-1, +1)$, that is, where the two attitudes are reversed, either complete rejection of concept 1 and complete acceptance of concept 2 occurs, or vice versa. These are the two peaks in Figure 2f. In a symmetric manner, in case the connection type, K_0 is positive, the vast majority of the population will either completely support both concepts (one of the peaks will be at $(+1, +1)$), or will completely reject both of them (the other peak will be at $(-1, -1)$), as in Figure 2c. That is, independently of the type of connection, the originally uniformly distributed attitudes will tend towards the extremities, meaning that the mere attempt to maintain a consistent belief system alone promotes the processing of attitudes tending towards extremities in case of being exposed to persistent information. Of course, in reality, it is not only one type of news that circulates within a community, but many types, often with different messages and connotations, but it is certainly an important—and so far overlooked—point, that this human drive (the urge to maintain consistent beliefs) alone has the capacity to push attitudes towards extremities—a phenomenon increasingly experienced in our increasingly connected world.
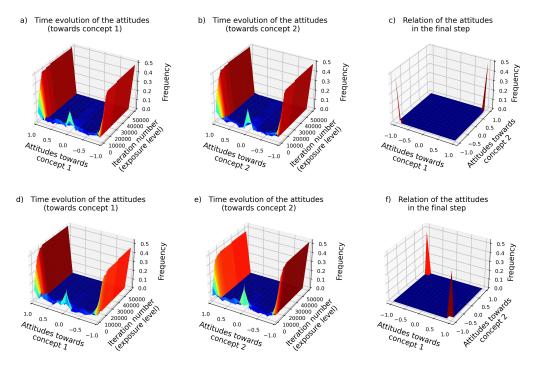

Figure 2. A typical time evolution of two attitudes ($a_{i,1}(t)$ and $a_{i,2}(t)$) within a population of $N = 100$ agents. Top row: $K_0 = +1$ (supportive relation). (**a**,**b**): the distribution of attitudes values towards concepts 1 and 2, respectively, as a function of time t. (**c**): At the final state, the vast majority either supports both beliefs (marked by the peak at $(1, 1)$) or rejects it (marked by the peak at $(-1, -1)$). Bottom row (**d–f**): $K_0 = -1$ (conflicting relation). (**f**) The major difference in this case is that at the end of the simulation most agents support one of the beliefs and disagree with the other (marked by the sharp peaks at the $(+1, -1)$ and $(-1, +1)$ points). The parameters are: population size $N = 100$, number of iterations $T = 50{,}000$, and connection type $K_0 = -1/+1$, and the noise value is $Z_A = 0.01$.

In Figure 3c "extremity" is defined as "being closer to +1 or −1 than a certain threshold value ϵ". Accordingly, if $\epsilon = 0.01$, then attitudes between 0.99 and 1, and attitudes between −0.99 and −1 will be considered as "extreme". Similarly, if $\epsilon = 0.1$, then attitudes between 0.9 and 1, and the ones between −0.9 and −1 will be considered as "extreme". Apparently, as can be seen in Figure 3c, the exact value of ϵ does not matter.

a) Time evolution of the attitudes
(towards concept 1)

b) Time evolution of the attitudes
(towards concept 1)
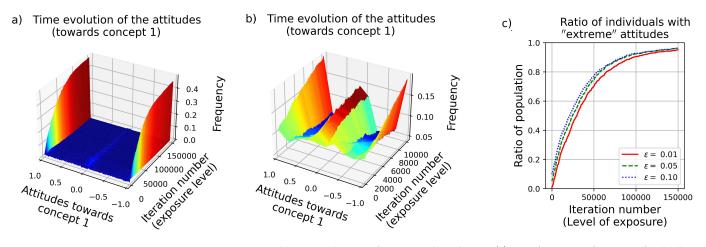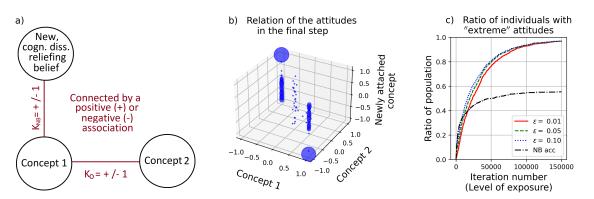
c) Ratio of individuals with
"extreme" attitudes

Figure 3. A typical time evolution of an attitude value $a_{i,k}(t)$ as a function of time (t) for (**a**) large population ($N = 1000$), and (**b**) under limited exposure to news ($T = 10{,}000$). In this case, since in each time step 1 individual learns the news (out of the $N = 1000$), on average, each agent will have heard it 10 times at the end of the simulation. As can be seen, for such a level of exposure, developing a neutral standpoint (adopting attitude values close to zero) is a good "strategy" as well. However, this neutrality vanishes in case of more enduring circulation of the news. (**c**) Proportional to the level of exposure (iteration number t), the ratio of the population holding "extreme" attitudes monotonically grows, independently of how "extremity" is defined (by the parameter ϵ). The parameters are: population size $N = 1000$, Number of iterations $T = 150{,}000$ (except for sub-figure (**b**), on which $T = 10{,}000$), connection type between the concepts $K_0 = -1$ and the noise value is $Z_A = 0.01$.

Note the small peaks around near-zero values in Figure 2a,b,d,e, at small t values. According to the simulations, *in case of limited exposure to the news*, agents might also adopt neutral standpoints (marked by near-zero attitude values) in order to avoid cognitive dissonance. This phenomenon is highlighted in Figure 3b. However, this is an unstable equilibrium point, since any further information regarding the given concept (appearing as noise Z_A in Equation (2)) pushes the attitude value away from zero. (See also Supplementary Information, Figure S3).

4.2. Finding Relief in New Ideas

As has already been mentioned, the other "basic strategy" applied by people in order to reduce the unpleasant feeling of cognitive dissonance is to reinterpret the incoming information by placing it into a context in which the contradiction vanishes, or even better, serves as a basis for reassurance [53]. For example, a doctor in his blog recollected memorable moments of the first year of the COVID-19 pandemic [69]. He remembers that when he tried to convince his family members to take the vaccine, he received vehement rejection, which was settled by receiving the comment that "You have good intentions, we know it. But you do not see the reality, because the "big players" leave you out from the party". As it turned out, by this they meant that the "big players" know perfectly well that the pandemic is a hoax, but they use the everyday doctors—such as the one writing the blog—for their purposes, i.e., to force "everyday people" into take the unnecessary and harmful vaccine. In this example, the doctor is a positive concept in the eye of his relatives, but the epidemic is negative (considered to be a hoax). When it turned out that the doctor considered the epidemic real (hence they should take the vaccine), he created a positive (supportive) relation between himself and the pandemic. This resulted in cognitive dissonance in the relatives, which was dissolved by adopting the new belief (about the "party" of the "big players"), which allowed the original attitudes to remain unchanged.

In the context of the present framework, this scenario can be represented by supplementing the original graph (including two nodes and an edge between them, as in Figure 1) with a new node, representing the new belief (see Figure 4a. The new belief can be related to either of the original concepts, or to both of them. As an example, in Figure 4a, the new

belief is connected to Concept 1. The type of connection, K_{NB}, can be either supportive or contradictory, similarly to the connection relating the two original concepts, K_0.

Figure 4. Change of attitudes within a population due to the circulation of some news, connecting concepts 1 and 2. In this case, agents *might* adopt a new belief as well, in case it reduces their cognitive dissonance. (**a**) The new belief can be connected to either or both concepts 1 and 2. (**b**) The "stable configuration" toward which the dynamics tends to (after 150,000 simulation steps). (**c**) The ratio of individuals holding "extreme beliefs", and adopting the new belief (black semi-dotted line). "Extreme attitudes" are those closer to +1 or −1 than a certain threshold value ϵ, such as 0.01, 0.05 and 0.1. As it can be seen, the ratio is largely independent of the exact value of ϵ. The parameters are: $N = 1000$, $T = 150{,}000$, $K_0 = -1$, $K_{NB} = -1$ and $Z_A = 0.01$.

Figure 4b depicts the "stable configuration" which the dynamics tends towards. As has been shown already, in the case of $K_0 = -1$, the attitudes towards concept 1 and 2 tend to be antagonistic and extreme (marked by the attitude values accumulating in the $(-1, +1)$ and $(+1, -1)$ points on the $x - y$ plain), while in the case of $K_0 = +1$ (positive relation), the attitudes towards concept 1 and 2 tend to be coincidental and also extreme (marked by the attitude values accumulating in the $(+1, +1)$ and $(-1, -1)$ points on the $x - y$ plain, see Supplementary Figure S5). The vertical, z axis depicts the attitude values towards the new, cognitive-dissonance-relieving belief; those who adopt it tend to develop an extreme relation towards this belief as well (in case of unceasing exposure). In contrast, those for whom the approval of the new belief would create cognitive dissonance, simply reject its adoption. In terms of the model, in their case, the edge K_{NB} will simply not exist, and hence the node representing this belief will not be connected to the belief network. The two solid columns belong to these agents, depicting their original attitudes, which simply do not change throughout the simulation. (In case we stipulate that only those values are shown in the figure which participate in the belief system of an agent, these values could be omitted as well, but for sake of clarity, in Figure 4b, we have kept them.) Furthermore, a small vertical "cloud" can be seen in the middle of the chart, representing those who are neutral towards the original concepts, i.e., their attitude towards the new (cognitive-dissonance-relieving) belief can take any value. Importantly, as is apparent from Figure 4, this mechanism pushes the attitudes towards extremities as well.

5. Discussion

The detailed methods by which humans perceive and make sense of the world—despite some eminent achievements [33,34,38,39]—is still to be understood. However, some basic characteristics have been elucidated by now, and have become part of mainstream science as well [32,35,36]. One such characteristic is that in the human mind, beliefs are strongly interconnected, and as such, no belief, concept or "piece of information" can exist on its own. Furthermore, in case of new information, humans immediately attempt to interlock it in a coherent way, seeking for connections and support with already existing beliefs.

Accordingly, the novelty of the present model lies not in "assuming" the above-mentioned two human characteristics—since they are well-studied and widely accepted by main-stream science [43,53,55]—rather, it lies in their mathematical formulation and incorporation into agent-based models.

There are two more further points worthy of consideration related to the model:

(i) Real belief systems have a tremendous amount of elements (instead of two or three), that are interconnected and embedded into each other in a complicated manner [39,43], and, accordingly, the "optimization process"—the attempt to minimize the contradictions among the components—refers to the *entire* system. From a physicist's point of view, this process is in close relation to physical structures aiming to reach an energy minimum. In this approach, "different realities" [57] can be different local energy minimums of similar systems. However, it is imperative to understand the *elementary relation* between *two* elements of the system before considering the entire structure. The present manuscript focuses on this elementary relation. Graph representation is important because, and only because, it serves as a mathematical tool for handling interrelated entities (which are the "beliefs" or "concepts" in our case). Since in the human mind a vast amount of concepts and beliefs are interrelated densely and intricately, any of its graph representations must also assume a vast amount of intricately interrelated (linked) nodes. However, from the viewpoint of the present study, the specific type of the graph does not play any role, because we focus on the elementary process altering the characteristics of two nodes (namely the "attitude values") due to a newly appearing link between them. (If a link appears, it is due to a certain piece of information connecting the two, originally unconnected beliefs/concepts). The nodes whose values alter are selected by the link (representing a piece of information).

(ii) The present model does not assume that the repeated information is *exactly* the same, only that the *type of connection* between two concepts (say a political party and a public issue, such as immigration or environmental topics) is tenaciously either positive or negative. Hence, it also explains how attitudes can become extreme due to the continuous repetition of information, and as such, it serves as a complementary explanation [70] for the reason why, throughout history, the most diverse regimes found it useful to repeat the same messages over and over again (despite the fact that everybody had already heard them many times).

Furthermore, the present model has some additional results as well, which are yet to be studied. Specifically, according to the results, the attitudes a certain type of news or information triggers depend on the *intensity* of the exposure. More precisely, in case of limited exposure, people tend to develop a centralist attitude first (which is an unstable equilibrium point), which, in case of persistent news-circulation, give way to extreme stances. The dynamics under *limited* exposure to news was not studied extensively in the present manuscript.

The ambition of the paper was to call attention to certain human traits that have not yet been incorporated into current computational models aiming to simulate opinion dynamics in human communities. From this perspective, the main point is the naturalness by which polarization can emerge, despite the fact that the model incorporates only minimal assumptions which are considered to be part of well-established, main-stream scientific results.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/e24091320/s1. Figure S1: Histogram of the attitude values, along with their standard deviations, for four time-steps: t = 0, 1000, 5000 and 30,000. Figure S2: Simulation results for $N = 100$ agents, that is, for one scale smaller population than the one studied in the main text. Figure S3: The evolution of attitudes within a population, due to limited exposure to news, achieved by limiting the number of iterations in $T = 3000$. Figure S4: Simulation results with four different noise values. Figure S5: The effect of the choice of K_0, the type of connection. Figure S6:

Simulation result with four different initial attitude value distributions. Figure S7: Incorporating the hierarchical nature of beliefs.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The source code of the simulation can be found on CoMSES, a public computational model library, under the following link: https://www.comses.net/codebases/2ccbae5 e-d51c-41ed-80c2-5e4b33c4e9c0/releases/1.0.0/ (Accessed on: 13 September 2022).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Flowchart and Parameters

The simulation was written in Python. Figure A1 shows the flowchart of the algorithm which is enough to replicate the results. However, the source code of the simulation can be found on CoMSES, a public computational model library as well [71]. Figure A1a shows the algorithm for the results explained in Section 4.1, and Figure A1b depicts the algorithm for the case when people might accept a new, cognitive-dissonance-relieving belief as well, the case explained in Section 4.2.
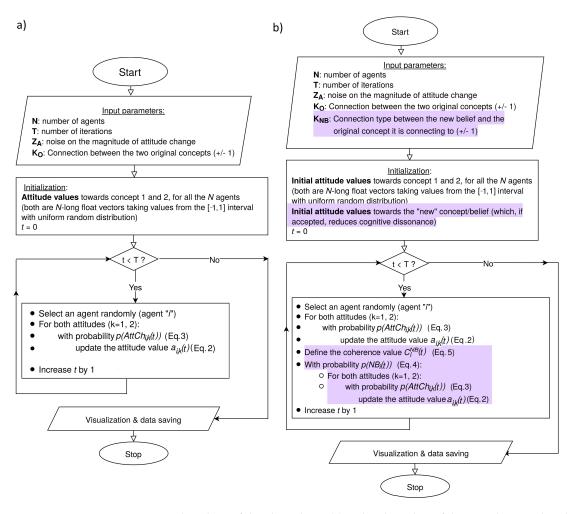

Figure A1. Flowchart of the algorithms. (**a**) is the algorithm of the case discussed in the Section 4.1. (**b**) depicts the algorithm detailed in the Section 4.2. The main difference—highlighted with purple—is that in the latter case, agents might adopt a new belief as well, in case it decreases their cognitive dissonance.

Since this is a minimalist model, altogether 5 parameters were used, which are overviewed in Table A1. The robustness of the results have been validated for a wide range of parameters, for which the results can be found in the Supplementary Material.

Table A1. Summary of the model parameters. Left column: Nomination used in the manuscript. Middle column: description, and right column: values used in the simulation.

Notation	Meaning	Values
N	Number of individuals (population size)	1000 (SI: 100)
T	Number of iterations (level of exposure)	100,000–300,000
Z_A	Noise on the magnitude of attitude change	$\in [-0.01, +0.01]$ (SI: $\in [-0.05, +0.05]$, $\in [-0.1, +0.1]$, $\in [-0.2, +0.2]$, $\in [-0.5, +0.5]$)
K_O	Positive or negative: Type of the connection between the two original concepts	+1 or −1
K_{NB}	Type of the connection between the newly accepted concept and the one it is connected to.	+1 or −1

References

1. Streitfeld, D. The Internet is broken. *The New York Times*, 20 May 2017.
2. Douven, I.; Hegselmann, R. Mis- and disinformation in a bounded confidence model. *Artif. Intell.* **2021**, *291*, 103415. [CrossRef]
3. Berekméri, E.; Derényi, I.; Zafeiris, A. Optimal structure of groups under exposure to fake news. *Appl. Netw. Sci.* **2019**, *4*, 101. [CrossRef]
4. Altmire, J. *Dead Center: How Political Polarization Divided America and What We Can Do About It*; Sunbury Press: Mechanicsburg, PA, USA, 2017.
5. Carothers, T.; O'Donohue, A. *Democracies Divided: The Global Challenge of Political Polarization*; Brookings Institution Press: Washington, DC, USA, 2019.
6. Ferguson, N. *The Square and the Tower: Networks and Power, from the Freemasons to Facebook*; Penguin Press: London, UK, 2018.
7. McPherson, M.; Smith-Lovin, L.; Cook, J.M. Birds of a Feather: Homophily in Social Networks. *Annu. Rev. Sociol.* **2001**, *27*, 415–444. Available online: http://xxx.lanl.gov/abs/doi:10.1146/annurev.soc.27.1.415 (accessed on 13 September 2022). [CrossRef]
8. Dandekar, P.; Goel, A.; Lee, D.T. Biased assimilation, homophily, and the dynamics of polarization. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5791–5796. Available online: http://xxx.lanl.gov/abs/https://www.pnas.org/content/110/15/5791.full.pdf (accessed on 13 September 2022). [CrossRef]
9. Perra, N.; Rocha, L.E.C. Modelling opinion dynamics in the age of algorithmic personalisation. *Sci. Rep.* **2019**, *9*, 7261. [CrossRef]
10. Berekméri, E.; Zafeiris, A. Optimal collective decision making: Consensus, accuracy and the effects of limited access to information. *Sci. Rep.* **2020**, *10*, 16997. [CrossRef]
11. Noorazar, H. Recent advances in opinion propagation dynamics: A 2020 Survey. *Eur. Phys. J. Plus* **2020**, *135*, 521. [CrossRef]
12. Lorenz, J. Continuous Opinion Dynamics under Bounded Confidence: A Survey. *Int. J. Mod. Phys. C* **2007**, *18*, 1819–1838. Available online: http://xxx.lanl.gov/abs/doi:10.1142/S0129183107011789 (accessed on 13 September 2022). [CrossRef]
13. Hegselmann, R.; Krause, U. Opinion dynamics and bounded confidence models, analysis, and simulation. *J. Artifical Soc. Soc. Simul. (JASSS)* **2002**, *5*, 1–33 .
14. Häggström, O.; Hirscher, T. Further results on consensus formation in the Deffuant model. *Electron. J. Probab.* **2014**, *19*, 26. [CrossRef]
15. Deffuant, G.; Neau, D.; Amblard, F.; Weisbuch, G. Mixing beliefs among interacting agents. *Adv. Complex Syst.* **2000**, *03*, 87–98. Available online: http://xxx.lanl.gov/abs/doi:10.1142/S0219525900000078 (accessed on 13 September 2022). [CrossRef]
16. Albi, G.; Pareschi, L.; Zanella, M. Boltzmann-type control of opinion consensus through leaders. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2014**, *372*, 20140138. [CrossRef] [PubMed]
17. Castellano, C.; Fortunato, S.; Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **2009**, *81*, 591–646. [CrossRef]
18. Mäs, M.; Flache, A. Differentiation without Distancing. Explaining Bi-Polarization of Opinions without Negative Influence. *PLoS ONE* **2013**, *8*, e74516. [CrossRef]
19. Edmonds, B. Assessing the Safety of (Numerical) Representation in Social Simulation. In Proceedings of the European Social Simulation Association (ESSA) 2005, Koblenz, Germany, 5–9 September 2005 .
20. Mäs, M.; Flache, A.; Helbing, D. Individualization as Driving Force of Clustering Phenomena in Humans. *PLoS Comput. Biol.* **2010**, *6*, e1000959. [CrossRef] [PubMed]

21. Pineda, M.; Toral, R.; Hernandez-Garcia, E. Noisy continuous-opinion dynamics. *J. Stat. Mech. Theory Exp.* **2009**, *2009*, P08001. [CrossRef]

22. Liu, Q.; Wang, X. Opinion dynamics with similarity-based random neighbors. *Sci. Rep.* **2013**, *3*, 2968. [CrossRef] [PubMed]

23. Schawe, H.; Hernández, L. When open mindeness hinders consensus. *Sci. Rep.* **2020**, *10*, 8273. [CrossRef]

24. Schawe, H.; Fontaine, S.; Hernández, L. When network bridges foster consensus. Bounded confidence models in networked societies. *Phys. Rev. Res.* **2021**, *3*, 023208. [CrossRef]

25. Berelson, B. Events as an Influence upon Public Opinion. *J. Q.* **1949**, *26*, 145–148. [CrossRef]

26. Atkeson, L.R.; Maestas, C.D., Extraordinary Events and Public Opinion. In *Catastrophic Politics: How Extraordinary Events Redefine Perceptions of Government*; Cambridge University Press: Cambridge, UK, 2012; pp. 1–21. [CrossRef]

27. Condie, S.A.; Condie, C.M. Stochastic events can explain sustained clustering and polarisation of opinions in social networks. *Sci. Rep.* **2021**, *11*, 1355. [CrossRef]

28. Toscani, G.; Tosin, A.; Zanella, M. Opinion modeling on social media and marketing aspects. *Phys. Rev. E* **2018**, *98*, 022315. [CrossRef] [PubMed]

29. Pareschi, L.; Toscani, G.; Tosin, A.; Zanella, M. Hydrodynamic Models of Preference Formation in Multi-agent Societies. *J. Nonlinear Sci.* **2019**, *29*, 2761–2796. [CrossRef]

30. Dalege, J.; Borsboom, D.; van Harreveld, F.; van den Berg, H.; Conner, M.; van der Maas, H.L.J. Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychol. Rev.* **2016**, *123*, 2–22. [CrossRef]

31. van der Maas, H.L.J.; Kolstein, R.; van der Pligt, J. Sudden Transitions in Attitudes. *Sociol. Methods Res.* **2003**, *32*, 125–152. [CrossRef]

32. Goldstein, E.B. *Cognitive Psychology: Connecting Mind, Research and Everyday Experience*, 5th ed.; Cengage Learning: Boston, MA, USA, 2018.

33. Seth, A. *Being You: A New Science of Consciousness*; Dutton: New York, NY, USA, 2021.

34. Pinker, S. *How the Mind Works*, Illustrated ed.; W. W. Norton & Company: New York, NY, USA, 2009.

35. Galotti, K.M. *Cognitive Psychology In and Out of the Laboratory*, Fifth ed.; SAGE Publications: New York, NY, USA, 2013.

36. Lamberts, K.; Shanks, D., Eds. *Knowledge Concepts and Categories*; Psychology Press: New York, NY, USA, 2016.

37. Harari, Y.N. *Sapiens: A Brief History of Humankind*; Harper: Manhattan, NY, USA, 2015.

38. Searle, J. *Making the Social World: The Structure of Human Civilization*; Oxford University Press: New York, NY, USA, 2010.

39. Shermer, M. *The Believing Brain: From Ghosts and Gods to Politics and Conspiracies—How We Construct Beliefs and Reinforce Them as Truths*, Illustrated ed.; St. Martin's Griffin: New York, NY, USA, 2012.

40. Pohl, R.F. (Ed.) *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*, 1st ed.; Psychology Press: London, UK, 2012.

41. Sloman, S.; Fernbach, P. *The Knowledge Illusion: Why We Never Think Alone*; Riverhead books: New York, NY, USA, 2017.

42. Cooper, J. *Cognitive Dissonance: 50 Years of a Classic Theory*; SAGE Publications Ltd.: New York, NY, USA, 2007.

43. Tóth, B.; Csányi, V. *Our Beliefs—The building blocks of human thoughts (In Hungarian: Hiedelmeink—Az Emberi Gondolatok építőkövei)*; Libri: Budapest, Hungary, 2017.

44. Kahneman, D. *Thinking, Fast and Slow*; Farrar, Straus and Giroux: New York, NY, USA, 2013.

45. March, J.G. *Primer on Decision Making: How Decisions Happen*; Free Press: New York, NY, USA, 2009.

46. Akerlof, G.A.; Kranton, R.E. *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*; Princeton University Press: Princeton, NJ, USA, 2010.

47. Converse, P.E. The nature of belief systems in mass publics. *Crit. Rev.* **1964**, *18*, 1–74. [CrossRef]

48. Jost, J.; Federico, C.; Napier, J. Political ideology: Its structure, functions, and elective affinities. *Annu. Rev. Psychol.* **2009**, *60*, 307–337. [CrossRef]

49. DellaPosta, D. Pluralistic Collapse: The "Oil Spill" Model of Mass Opinion Polarization. *Am. Sociol. Rev.* **2020**, *85*, 507–536. [CrossRef]

50. Turner-Zwinkels, F.M.; Johnson, B.B.; Sibley, C.G.; Brandt, M.J. Conservatives' Moral Foundations Are More Densely Connected Than Liberals' Moral Foundations. *Personal. Soc. Psychol. Bull.* **2020**, *47*, 167–184. [CrossRef] [PubMed]

51. Rokeach, M. The organization and modification of beliefs. *Centen. Rev.* **1963**, *7*, 375–395.

52. Goldberg, A.; Stein, S.K. Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation. *Am. Sociol. Rev.* **2018**, *83*, 897–932. [CrossRef]

53. Festinger, L. *A Theory of Cognitive Dissonance*; Stanford University Press: Redwood City, CA, USA, 1957.

54. Leonard, J. Cognitive Dissonance: What to Know. 2019. Available online: https://www.medicalnewstoday.com/articles/326738 (accessed on 13 September 2022).

55. Harmon-Jones, E. (Ed.) *Cognitive Dissonance: Reexamining a Pivotal Theory in Psychology*, 2nd ed.; American Psychological Association: Washington, DC, USA, 2019.

56. Perlovsky, L. A challenge to human evolution—Cognitive dissonance. *Front. Psychol.* **2013**, *4*, 179. [CrossRef] [PubMed]

57. Galef, J. *The Scout Mindset: Why Some People See Things Clearly and Others Don't*; Portfolio: London, UK, 2021.

58. Zafeiris, A.; Vicsek, T. *Why We Live in Hierarchies: A Quantitative Treatise*; Springer: Cham, Switzerland, 2018.

59. Brandt, M.J.; Sleegers, W.W.A. Evaluating Belief System Networks as a Theory of Political Belief System Dynamics. *Personal. Soc. Psychol. Rev.* **2021**, *25*, 159–185. [CrossRef] [PubMed]

60. Wright, R. *The Evolution of God*; Little, Brown and Company: New York, NY, USA, 2009.
61. Christian, D. *This Fleeting World: A Short History of Humanity*; Berkshire Publishing Group: Great Barrington, MA, USA, 2014.
62. Rodriguez, N.; Bollen, J.; Ahn, Y.Y. Collective Dynamics of Belief Evolution under Cognitive Coherence and Social Conformity. *PLoS ONE* **2016**, *11*, e0165910. [CrossRef]
63. Baldassarri, D.; Goldberg, A. Neither Ideologues nor Agnostics: Alternative Voters' Belief System in an Age of Partisan Politics. *Am. J. Sociol.* **2014**, *120*, 45–95. [CrossRef]
64. Brandt, M.J.; Sibley, C.G.; Osborne, D. What Is Central to Political Belief System Networks? *Personal. Soc. Psychol. Bull.* **2019**, *45*, 1352–1364. [CrossRef]
65. Priest, H. *Biased: 50 Powerful Cognitive Biases That Impair Our Judgment*; Cognitt Consulting: Vilnius, Lithuania, 2016.
66. Holm, C. *The 25 Cognitive Biases: Uncovering The Myth Of Rational Thinking*, Kindle ed.; 2015 .
67. Dobelli, R. *The Art of Thinking Clearly*; Harper Paperbacks: New York, NY, USA, 2014.
68. George, J.M.; Jones, G.R. Towards a process model of individual change in organizations. *Hum. Relations* **2001**, *54*, 419–444. [CrossRef]
69. It Is Not Condolence, Applaud or Awards That We Wish for—The Confessions of a Doctor about the Tribulations of the Covid Pandemic. Available online: https://miapanasz.reblog.hu/nem-sajnalatra-tapsra-vagy-kituntetesre-vagyunk (accessed on 7 January 2012). (In Hungarian)
70. Centola, D. *How Behavior Spreads: The Science of Complex Contagions*; Princeton University Press: Princeton, NJ, USA, 2018.
71. Zafeiris, A. Source Code of the Simulation on CoMSES Computational Model Library. 2022. Available online: https://www.comses.net/codebases/2ccbae5e-d51c-41ed-80c2-5e4b33c4e9c0/releases/1.0.0/ (accessed on 13 September 2022).