

Article

Modeling Misinformation Spread in a Bounded Confidence Model: A Simulation Study

Yujia Wu * and Peng Guo

School of Management, Northwestern Polytechnical University, Xi'an 710021, China; guopeng@nwpu.edu.cn

* Correspondence: wuyujia@mail.nwpu.edu.cn

Abstract: Misinformation has posed significant threats to all aspects of people's lives. One of the most active areas of research in misinformation examines how individuals are misinformed. In this paper, we study how and to what extent agents are misinformed in an extended bounded confidence model, which consists of three parts: (i) online selective neighbors whose opinions differ from their own but not by more than a certain confidence level; (ii) offline neighbors, in a Watts–Strogatz small-world network, whom an agent has to communicate with even though their opinions are far different from their own; and (iii) a Bayesian analysis. Furthermore, we introduce two types of epistemically irresponsible agents: agents who hide their honest opinions and focus on disseminating misinformation and agents who ignore the messages received and follow the crowd mindlessly. Simulations show that, in an environment with only online selective neighbors, the misinforming is more successful with broader confidence intervals. Having offline neighbors contributes to being cautious of misinformation, while employing a Bayesian analysis helps in discovering the truth. Moreover, the agents who are only willing to listen to the majority, regardless of the truth, unwittingly help to bring about the success of misinformation attempts, and they themselves are, of course, misled to a greater extent.

Keywords: misinformation; bounded confidence; opinion dynamics; small-world networks; heterogeneity



Citation: Wu, Y.; Guo, P. Modeling Misinformation Spread in a Bounded Confidence Model: A Simulation Study. *Entropy* **2024**, *26*, 99. <https://doi.org/10.3390/e26020099>

Academic Editor: Francisco W. De Sousa Lima

Received: 13 December 2023

Revised: 21 January 2024

Accepted: 22 January 2024

Published: 23 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Misinformation has become an inextricable part of the modern world, and the widespread dissemination of which has posed significant threats to all areas of people's lives. During the COVID-19 pandemic, the exposure to misinformation prompted engagement in misinformed behaviors and discouraged evidence-based prevention behaviors [1,2]. Misinformation about climate change confuses the public, leads to political inaction, and leads to the rejection of mitigation policies [3].

The spread of misinformation on social networks has been the central focus of some recent work. Some studies consider binary opinion dynamics. For example, the topic might be believing or not believing a certain message. Acemoglu and Mostagir considered whether agents will effectively weed out incorrect beliefs according to dispersed information and thus learn the truth [4,5]. Nguyen et al. focused on how many agents will be influenced by misinformation to take incorrect actions and how to minimize the number of those agents [6–8]. Besides taking two extremes (true or false), one can envision a continuous opinion space but the actions taken are discrete [9–11]. Liu et al. inferred the diffusion rate of misinformation and whether agents can debunk misinformation by calculating the densities of users with different opinions [12,13].

Unlike the limited options presented in those papers, the opinions can sometimes be represented by continuous variables. A particular type of agent holds an opinion that can be any real number in the interval $[0, 1]$, for example, 0.8. Researchers run simulations to determine the average number of agents that started with different initial opinions and have

been converted to a fixed point (i.e., 0.8). Most of this work assumes that the social network consists of strictly sincere agents, all willing to tell what they really think when communicating with others [14–17]. Douven and Hegselmann considered the possibility that some agents may have ulterior motives, including carrying out misinformation campaigns [18]. In this study, the agents' opinions are represented by continuous variables in the interval $[0, 1]$, where zero and one illustrate the misinformation and the truth, respectively. We investigated how individuals are misled to the misinformation side.

Much of the literature studying continuous opinions focuses on the bounded confidence model in which an agent only interacts with those whose opinions differ from their own by less than a certain value. Two well-known bounded confidence models were introduced by Hegselmann and Krause (HK) [19–21] and Deffuant, Weisbuch et al. (DW) [22,23]. These two models differ in the communication regime: in the DW model, agents meet in random pairwise encounters and change their opinions to the average of both opinions if they are within each other's confidence intervals; in the HK model, the most well-known synchronous version of the DW model, each agent takes the arithmetic mean of opinions within their confidence interval, and all agents update at the same time.

There are many extensions of the bounded confidence model in different respects. Some works assume situations where all individuals have the same level of confidence [24–26]. However, more extensions assume that the agents' bounds of confidence are heterogeneous. Lorenz proposed a society of agents with two different bounds of confidence (open-minded with a broader confidence interval and closed-minded with a narrower one) [27,28]. The confidence radii in some works are even agent-specific [29–34] or time-varying [18,35]. Furthermore, Hegselmann et al. considered asymmetric confidence intervals, where an agent took a broader range of opinions on their left (right) than on their right (left) into consideration if we represented the opinions on an x-axis [15,21,26,35]. Some researchers also considered agents who never update their opinions and examine their impact on other normal individuals and the trend of public opinion in social networks. Some papers refer to such agents as opinion leaders [36,37], others as mass media [33,38], and a few as stubborn agents [39,40].

Inspired by those studies, we present an extended BC model with "stubborn agents" and heterogeneous bounds of confidence. What makes our model different from the previous ones is that the agents adopt different confidence intervals for different kinds of neighbors. Knop et al. compared online and offline communication and argue that, though people exchange messages more often online, face-to-face communication leads to a higher amount and more depth of self-disclosure [41]. Thus, we bring not only online but also offline opinion dynamics into the model, where the agents act open-minded when facing their offline neighbors but are relatively closed-minded when interacting with their online selective neighbors. Furthermore, "stubborn agents" appear in the form of "paid posters", a type of epistemically irresponsible agent, who deliberately disseminate misinformation and keep their opinions unchanged at "0". An interesting and practical assumption is that paid posters are only online virtual agents created by some normal individuals.

The averaging process mentioned above is also a basic approach usually used in non-Bayesian methods. A large amount of literature has studied how misinformation affects social learning in two canonical models: Bayesian and non-Bayesian learning. In these models, agents try to learn the truth by exchanging opinions with each other and learning from the messages received from the news. Non-Bayesian agents naively update beliefs by repeatedly taking weighted averages of their neighbors' opinions [42]. In contrast, based on an underlying model of the reality, Bayesian agents are presumed to update their beliefs optimally (from a statistical perspective). Bayesian approaches require agents to have a reliable understanding of the world for assigning priors and to take others' strategies into account [43]. Although these requirements make issues such as indoctrination and the spread of misinformation nearly impossible, they are still quite challenging [44]. Therefore, besides aggregating the opinions of the neighbors in a non-Bayesian way, we assume that

an agent also computes their posterior beliefs by applying a simplified Bayesian analysis conditional on the messages received about the truth.

Furthermore, we introduce another type of epistemically irresponsible agent: agents who hide their honest opinions and focus on disseminating misinformation and who ignore the messages received and follow the crowd mindlessly.

The simulations illustrate that in an environment with only online selective neighbors, misinforming is more successful with broader confidence intervals. Having offline neighbors contributes to being cautious of misinformation, while employing Bayesian analysis helps in discovering the truth. Moreover, the agents, who are only willing to listen to the majority, whatever the truth, unwittingly help to bring about the success of misinformation attempts, and they themselves are, of course, misled to a greater extent.

We next provide a brief outline of the rest of this paper. In Section 2, we present an opinion update model with bounded confidence intervals and heterogeneous agents. Section 3 discusses the simulations and analyses. In Section 4, we conclude the paper.

2. Model

2.1. The Original BC Model

There is a true state of the world θ (whose exact nature we do not specify here). For simplicity, we assume that the opinion concerning θ can be expressed by a real number. Labeled by $i \in V = \{1, 2, \dots, n\}$, n agents interact with each other in a small social network at discrete points in time. Agent i 's opinion concerning θ at time t is denoted by $\mu_{i,t} \in [0, 1]$. The two endpoints of that closed interval represent two extremely opposite attitudes towards θ : $\mu_i = 1$ when agent i considers θ to be true and $\mu_i = 0$ when agent i is completely misinformed by misinformation and considers θ to be false. As for the real numbers between those two endpoints, agents with opinions $\mu_{i,t} \geq 0.5$ are inclined to believe θ , whereas those with opinions $\mu_{i,t} < 0.5$ are more attracted to misinformation. Therefore, μ_i can also be interpreted as the probability that agent i believes θ .

In the BC model, when agents update their opinions, they only take into account those agents who are within their bounded confidence interval, meaning that their opinions differ from the agents less than a certain confidence level r_i . The set of agents within agent i 's bounded confidence interval at time t is given by

$$S(i, \mu_t) = \{1 \leq j \leq n \mid |\mu_{i,t} - \mu_{j,t}| \leq r_i\}$$

Thus, agent i 's opinion concerning θ at time $t + 1$ is defined to be

$$\mu_{i,t+1} = |S(i, \mu_t)|^{-1} \sum_{j \in S(i, \mu_t)} \mu_{j,t} \quad (1)$$

where $|\cdot|$ denotes the number of elements.

2.2. The Opinion Update Model

Two communication patterns coexist in the social network: online and offline communication. When online, agents only communicate with those who fall within their bounded confidence intervals. In particular, anyone can be anyone's neighbor regardless of their identities in real life. The aforementioned set of agents $S(i, \mu_t)$ will henceforth be called "online selective neighbors". Offline neighbors are people whom an agent has to communicate with, although their opinions are far different from their own, and who may deeply influence them, such as their families and colleagues. Since we would like to discuss a network with ordinary non-media agents who have a similar number of neighbors, we represent this offline social network with a Watts–Strogatz small-world graph $G(V_n, E)$, where each node in the set V_n represents an agent and each undirected edge in the set E represents the connection between two agents (see Figure 1). Specifically, we say that agent i and agent j know each other if there is an undirected edge between i and j . Given that

offline connections are relatively more stable, we assume that offline neighbors, denoted by $N(i, \mu_t)$, do not change in time.

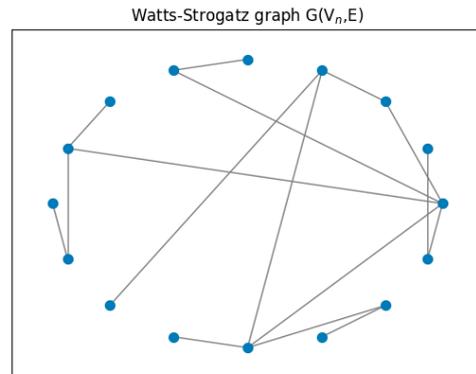


Figure 1. A Watts–Strogatz graph $G(V_n, E)$ where each node in the set V_n represents an agent and each undirected edge in the set E represents the connection between two agents.

To not complicate things too much, we assume that agent i equally weights all $j \in N(i, \mu_t)$:

$$\mu_{i,t+1}' = |N(i, \mu_t)|^{-1} \sum_{j \in N(i, \mu_t)} \mu_{j,t}. \tag{2}$$

Here, we add one more component to the opinion update process: “Bayesian analysis”. It has three key components: (i) priors, which are the agents’ existing beliefs about the underlying model of the world, and in this case, they are the agents’ opinions concerning θ at the previous time step, i.e., $\mu_{i,t}$; (ii) sources of information, which are the messages about θ that the agents receive from their surroundings or the news; and (iii) a method of information processing, i.e., Bayes’ rule. As a result, agent i forms a posterior opinion:

$$\mu_{i,t+1}'' = \frac{(1 - q)\mu_{i,t}}{(1 - q)\mu_{i,t} + q(1 - \mu_{i,t})}. \tag{3}$$

where q denotes the probability that a message contains misinformation. In particular, $q < 0.5$.

Thus, by combining Equations (1)–(3), agent i updates their opinions according to:

$$\mu_{i,t+1} = \alpha |S(i, \mu_t)|^{-1} \sum_{j \in S(i, \mu_t)} \mu_{j,t} + \beta |N(i, \mu_t)|^{-1} \sum_{j \in N(i, \mu_t)} \mu_{j,t} + \gamma \frac{(1 - p)\mu_{i,t}}{(1 - p)\mu_{i,t} + p(1 - \mu_{i,t})} \tag{4}$$

for weights α , β , and γ with $\alpha + \beta + \gamma = 1$.

2.3. Typed Agents

The agents that populate Equation (4) will henceforth be called “normals”. Now, we introduce two new types of agents. The first is “paid posters”, who hide their honest opinions and focus on disseminating misinformation. They always stick to a fixed opinion (zero) and never update. It is worth noting that the behavior mentioned above is only applicable to online platforms. We can think of “paid posting” as a job. So, in one sense, paid posters are only online virtual agents created by some normals. In particular, there is not a one-to-one correspondence between normals and paid posters, meaning that one normal can create as many paid posters as she wants. As shown in Figure 2, normal 01 and normal 03 are only normals. They behave the same online and offline. Normal 02 accepts the job of “paid posting” and creates four virtual agents, i.e., paid posters 01 to 04. As a result, in addition to expressing their true opinions on online platforms, normal 02 also disseminates misinformation as paid posters.

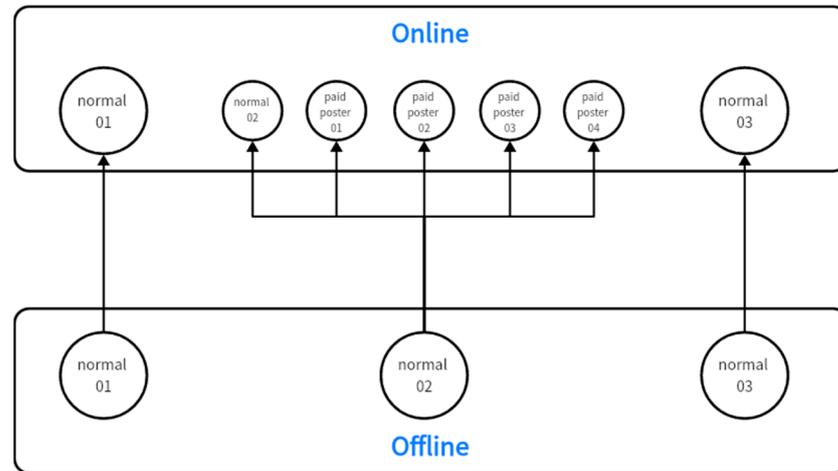


Figure 2. Besides being themselves, normal 02 also creates 4 paid posters online.

The second new type of agent is “fence-sitters”, who ignore the messages about θ and follow the crowd mindlessly. In Section 2.1, we divided agents into two groups: agents with opinions $\mu_{i,t} \geq 0.5$ that are inclined to believe θ and those with opinions $\mu_{i,t} < 0.5$ that are more attracted to misinformation. Thus, 0.5 is a boundary between two sides, and the majority is the side with the larger number of agents. Thus, whatever the truth, fence-sitters are only willing to listen to the majority:

$$\mu_{i,t+1} = |I(i, \mu_t)|^{-1} \sum_{j \in I(i, \mu_t)} \mu_{j,t} \tag{5}$$

where $I(i, \mu_t)$ denotes the set of agents on the majority side ($\mu_t < 0.5$ or $\mu_t \geq 0.5$).

Then, the formal opinion update model is as follows:

$$\mu_{i,t+1} = \begin{cases} \alpha |S(i, \mu_t)|^{-1} \sum_{j \in S(i, \mu_t)} \mu_{j,t} + \beta |N(i, \mu_t)|^{-1} \sum_{j \in N(i, \mu_t)} \mu_{j,t} \\ \quad + \gamma \frac{(1-q)\mu_{i,t}}{(1-q)\mu_{i,t} + q(1-\mu_{i,t})} & \text{if } i \text{ is a normal agent} \\ |I(i, \mu_t)|^{-1} \sum_{j \in I(i, \mu_t)} \mu_{j,t} & \text{if } i \text{ is a fence-sitter} \\ 0 & \text{if } i \text{ is a paid poster} \end{cases} \tag{6}$$

3. Simulations and Analyses

3.1. Online Selective Neighbors Only

Here, we consider the baseline case studied throughout the BC model literature. There are only selected neighbors and no offline neighbors or Bayesian analysis, that is, $\alpha = 1$, $\beta = 0$, and $\gamma = 0$. The social network we utilized here and elsewhere include 50 normals whose initial opinions are drawn randomly from the interval $[0, 1]$. We assume that the confidence level is the same for all agents, i.e., $\forall i \in V, r_i = r$. Besides exploring the number of agents who are inclined to believe θ ($\mu_{i,t} \geq 0.5$) or believe the misinformation ($\mu_{i,t} < 0.5$), we paid more attention to the agents who have been completely misinformed. Theoretically, an agent with an opinion $\mu_i = 0$ can be considered completely misinformed. However, to reduce the number of calculations, we say agent i completely believes the misinformation if she holds an opinion that almost equals zero, i.e., $\mu_i < 10^{-2}$.

In the first experiment, we considered a variable number of paid posters and different values of r . In this setup, the questions that concerned us were how many normals would be completely misinformed and to what extent this depends on the number of paid posters and the value of r present in the social network.

Figure 3 shows, for each combination of the number of paid posters (from 10 to 50, in steps of 5) and the value of r (from 0.1 to 0.4, in steps of 0.1), the number of normals whose

opinions have dropped below 10^{-2} in two simulations. For each of these nine steps of four different values of r , we repeat the simulation 50 times, starting with a different random start distribution each time. Each simulation was carried out until the dynamics became stable (stable means that for all agents i , $|\mu_{i,t+1} - \mu_{i,t}| \leq 10^{-5}$).

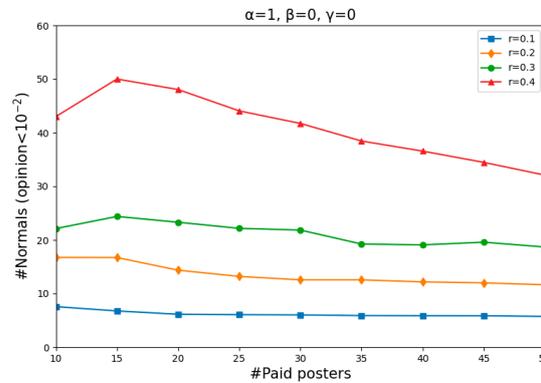


Figure 3. The number of normals whose opinions dropped below 10^{-2} , with the number of paid posters increasing from 10 to 50 (in increments of 5) and r rising from 0.1 to 0.4 (in increments of 0.1). “#Paid posters” denotes the number of paid posters and “#Normals” denotes the number of normals.

As shown in Figure 3, how many normals become entirely misinformed, on average, mainly depends on the willingness to count others as their peers. Specifically, the larger the value of r , the more successful the misinforming was. However, the results are largely insensitive to the number of paid posters when r is no more than 0.3 as there was not much difference in the number of normals who became entirely misinformed when there were 10 and 50 paid posters.

An obvious decline occurred with an increase in the number of paid posters when $r = 0.4$, meaning that the misinforming is more successful with fewer paid posters than with more of them, even though one might have expected to find the opposite. Such a puzzling phenomenon is actually quite understandable. Two single runs shown in Figure 4 illustrate the dynamics of $r = 0.4$ in more detail, one featuring 15 paid posters (left panel) and the other 50 (right panel). Since the starting distribution is not uniform, less populated areas, namely “gaps”, occur on the vertical lines ($t = 0$). The more paid posters there are, the harder they pull normals to zero, and the more likely a divergence occurs when there is a “gap”. In that case, the normals above the “gap” will never believe the misinformation. In contrast, fewer paid posters with a relatively weaker pull may attract all normals step by step and eventually convince them.

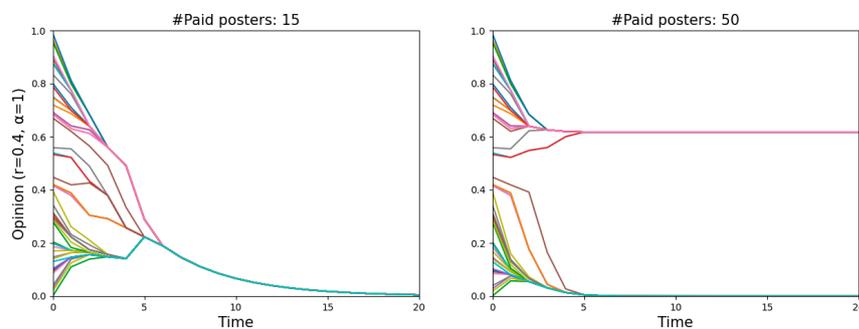


Figure 4. Illustration of why the misinforming is more successful with fewer paid posters than with more of them.

3.2. Adding Offline Neighbors and Bayesian Analysis

Now, let us look at a more interesting setting: $\alpha = 0.5$ and $\beta = 0.5$, indicating that normals update their opinions considering not only online selective neighbors but also

offline neighbors. We assume a Watts–Strogatz small-world network as the offline network. Since the social network we present is a small one, we started with a ring lattice of 50 vertices with two edges per vertex, and each edge is then rewired at random with a probability of 0.5. Figure 5 gives an example of such a network.

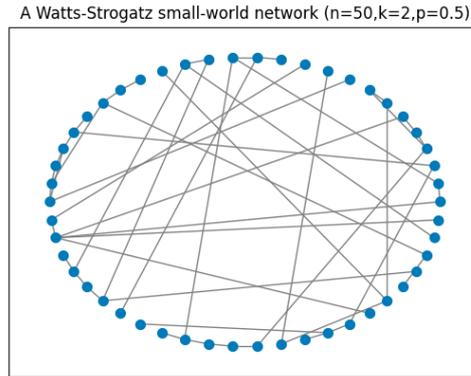


Figure 5. An example of a Watts–Strogatz small-world network with $n = 50$, $k = 2$, and $p = 0.5$.

What Figure 3 and the left panel of Figure 6 have in common is that the larger the value of r , the more successful the misinforming was. As can be seen in the left panel of Figure 6, when $r = 0.1$ and $r = 0.2$, even though there were many paid posters, almost no one completely believed the misinformation. Because offline neighbors are usually families and colleagues who hide their virtual features in real life and communicate with each other more sincerely, normals will receive much less deliberate misinformation. That is to say, offline communication enhances the normals’ abilities to beware of misinformation when normals are relatively closed-minded. However, when $r = 0.3$ and $r = 0.4$, the opposite situation occurs. The two ascending curves mean that the misinforming was more successful with more paid posters than with fewer of them. This is because the “gaps” that divided the normals into two camps in Figure 3 did not matter anymore here. Agent i ’s offline neighbors, whose opinions might differ significantly from agent i ’s, can pull them to any side of the gap. Thus, more paid posters have a higher chance of pulling normals to “opinion zero”.

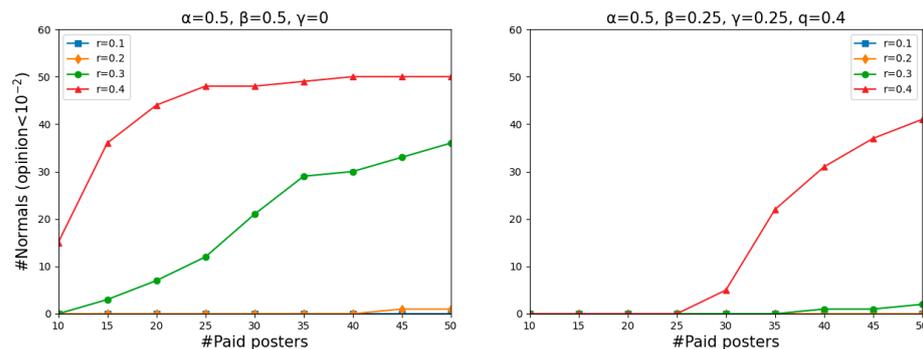


Figure 6. The number of normals, on average, whose opinions have dropped below 10^{-2} after adding offline neighbors (left panel) and further adding Bayesian analysis (right panel).

We found that all normals can always reach a consensus after adding offline neighbors, signifying that either all normals or none became entirely misinformed (see Figure 7). Since the numbers shown in Figure 6 are averaged over 50 simulations (starting with a different random start distribution each time), many agents’ opinions dropping below 10^{-2} can be interpreted as a high probability of all normals being completely misled.

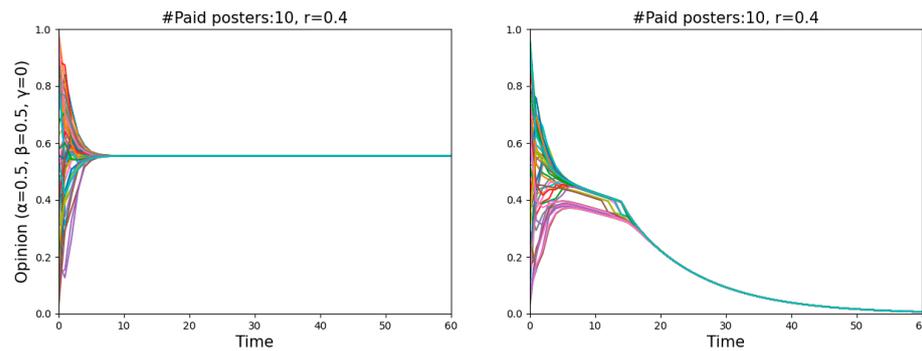


Figure 7. Two runs with identical parameter settings illustrate that for different starting distributions, normals sometimes will, and sometimes will not, become entirely misinformed.

The right panel of Figure 6 shows the results after further adding a Bayesian analysis. Recall that the q is the probability that a message contains misinformation. We only consider the case where there is much misinformation and set q to be 0.4. Agents are more capable of being cautious of misinformation, where almost no one completely believes the misinformation even when r increases to 0.3. When $r = 0.4$, only a few normals were entirely misinformed when there were fewer paid posters. With the number of paid posters increasing from 30, the number of normals being pulled to the opposite of the truth moved upward rapidly. The changes in numbers may at first appear sudden, but in fact, they are easy to explain. The combination of a large q and a few paid posters indicates that a small number of agents are sending the same messages repeatedly. When there is rampant misinformation but a handful of people sending it, normals can employ their sophisticated reasoning abilities to infer from this paradox that some agents are spreading rumors and choose not to believe. When there are as many misleading messages as paid posters, normals will relax their vigilance to stand on the side of the misinformation that more and more people are disseminating because of the herding effect. As can be seen, agents, even with the Bayesian analysis, are much easier to be misled when there are many paid posters and a large amount of misinformation.

Focusing on the point “#paid posters: 50, $r = 0.1$ ”, the results in the two panels of Figure 6 are the same. However, on closer inspection, as seen in Figure 8, agents are near the dividing line between believing and disbelieving the misinformation in the left panel, and agents reach a consensus to believe the truth in the right panel. To put it differently, having offline neighbors contributes to being cautious of misinformation, while employing a Bayesian analysis helps in discovering the truth.

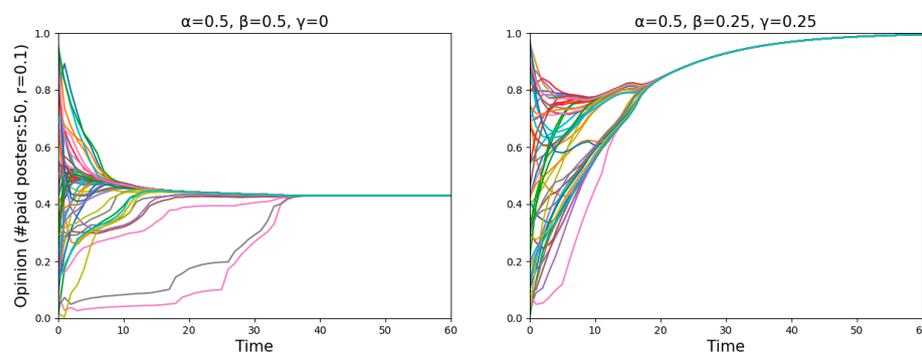


Figure 8. Comparison of two runs with identical parameter settings after adding offline neighbors (left panel) and further adding Bayesian analysis (right panel).

3.3. Adding Fence-Sitters

From this point on, we will only discuss the results simulated by the formal opinion update model in Section 2.3 (Equation (6)), which adds offline neighbors and a Bayesian analysis to the original BC model.

Until now, we have focused on social networks that consist of normals and paid posters. In this section, we consider the opinion dynamics with the addition of fence-sitters, who ignore the messages about θ and follow the crowd mindlessly. More exactly, we adopt six settings whose results are visualized in Figure 9: the top row with 10 paid posters, the middle row with 30, and the bottom row with 50. Each row contains two parts: the left column shows the results for opinions that dropped below 10^{-2} and the right column shows the results for opinions that were less than 0.5. To reduce complexity, we assumed that, for every fence-sitter, the number of the elements in their set of neighbors was 20, which comes from rounding the average of the numbers of all normals' online selective neighbors at the starting points in 20 simulations ($r = 0.2$). The question we were interested in is what effect the presence of different numbers of fence-sitters has on the number of normals who have been misinformed.

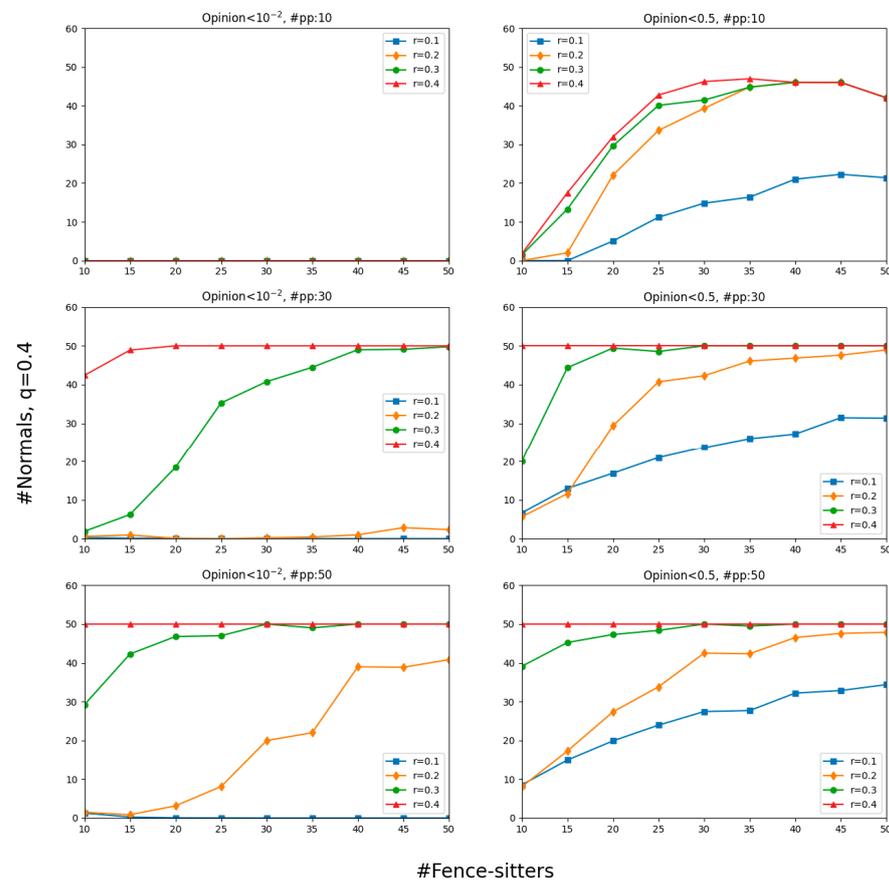


Figure 9. Under conditions of varying numbers of paid posters (top row: 10; middle row: 30; and bottom row: 50), the normals believe the misinformation to different degrees. The left column shows the results for the number of normal with opinions dropping below 10^{-2} and the right column shows the results for opinions less than 0.5. Results are averages over 50 simulations per number of fence-sitters and value of r combination. “#pp” is an abbreviation of “#paid posters”.

As mentioned before, normals are not easily misled by misinformation when there are a few paid posters. However, many normals were misinformed with the gradual increase in the number of fence-sitters. In other words, the rising lines in Figure 9 indicate that the more fence-sitters there are, the more normals will lower their opinions to less than 0.5. In particular, the number of misguided normals with 50 fence-sitters surpassed that without

fence-sitters—by far in most cases. That number, however, appears to be insensitive to the number of paid posters when there were more than 30: the risk brought about by “#pp: 50” was only slightly greater than that brought about by “#pp: 30”.

No matter how many fence-sitters there were and how liberal the normals were in counting others as their peers, no normals completely believed the misinformation (see upper left panel of Figure 9). Would this signify that the fence-sitters here were ineffective? The answer is no. The upper right panel of Figure 9 shows that more and more normals were pulled to the side of misinformation with a gradual increase in the number of fence-sitters. They begin to doubt the truth and gravitate towards the misinformation. The comparison between the two columns in Figure 9 also suggests that fence-sitters made more efforts to attract normals to the wrong side than to make them thoroughly misinformed.

Figure 10 shows the points “#fence-sitters:20, $r = 0.1$ ” and “#fence-sitters:50, $r = 0.1$ ” in the first row of Figure 9 in more detail. The fence-sitters’ opinions moved downward rapidly because of the existence of a few paid posters. The sudden increase in the number of agents who were inclined to believe the misinformation reinforced the fence-sitters’ conviction that they had made the correct choice. Vacillating between other fence-sitters’ opinions and zero, they never converged. Because the fence-sitters, as unwitting accomplices of paid posters, could appear in any normal agent’s set of offline neighbors, some normals may find the misinformation attractive and lower their values of opinions. Consequently, the presence of fence-sitters had an obvious negative effect on the normals. It only took a few fence-sitters to make the normals who have already reached a consensus on the truth appear indecisive about believing it or not.

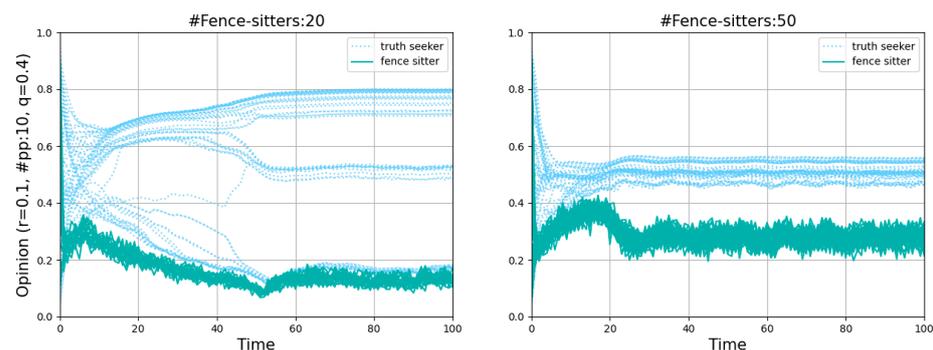


Figure 10. Comparison of opinion dynamics with 20 and 50 fence-sitters where there are a few paid posters and a small value of r .

Fence-sitters vacillated between other fence-sitters’ opinions and zero, and their opinions stayed between 0 and 0.5. Their exact position between 0 and 0.5 depended on the values of other parameters. The more misinformation there was, the more successful the paid posters were with the help of the fence-sitters, and the more the opinions of normals moved downward. Once the opinions of normals declined to below 0.5, the fence-sitters were then pulled up to be near the normals. Then, normals maintained their opinion, and the fence-sitters continued to vacillate between the normals and zero. Consequently, except for very few normals whose opinions were the nearest to zero at the very beginning, all of them were further from zero than the fence-sitters were.

As for the more general lesson we have learned from Figure 9, fence-sitters unwittingly help paid posters to bring about the success of misinformation attempts, but not to the point of making normals completely misinformed. Fence-sitters themselves are, of course, misled and nearer to zero than normals are.

4. Conclusions

Our project started from the assumption that the widespread misinformation in social networks, disseminated by ill-disposed agents, can mislead agents. Since rampant misinformation has posed significant threats to all areas of people’s lives, there is a practical

interest in examining how and to what extent agents are misinformed by misinformation. As long as we understand where our greatest vulnerabilities lie when communicating with others, we are able to take measures to protect ourselves against epistemic manipulation.

With this in mind, we revisited the model, which combines Bayesian and non-Bayesian learning, and assumed the network topology to be a Watts–Strogatz small-world network. Therefore, we first proposed an extension of the model in which neighbors are divided into two parts: online selective neighbors whose opinions differ from their own but not by more than a certain confidence level, and offline neighbors whom an agent has to communicate with even though their opinions are far different from their own. In a further step, we introduced two new types of agents: first, paid posters, who only exist online, always stick to a fixed opinion (zero), and never update; second, fence-sitters, who are only willing to listen to the majority, regardless of the truth.

In an environment with only selective neighbors, the misinforming is more successful with broader confidence intervals. Having offline neighbors contributes to being cautious of misinformation, while employing a Bayesian analysis helps in discovering the truth. However, a Bayesian analysis does not work when there are many paid posters sending misinformation. Moreover, fence-sitters unwittingly help paid posters to bring about the success of misinformation attempts, but not to the point of making normals completely misinformed. Fence-sitters themselves are, of course, misled and nearer to zero than normals are.

We find that agents on the majority side serve as stochastic exogenous signals to make fence-sitters carry out random jumps inside the whole opinion space [45]. Influenced by those agents, fence-sitters are always vacillating about whether to believe the misinformation. Therefore, exploring more interesting phenomena induced by stochastic exogenous signals is left to be a future project.

Our research is subject to some limitations. Since this is a simulation study, we hope to conduct an empirical study based on real-world data to calibrate the model in the future. Furthermore, we represented the offline social network with a Watts–Strogatz small-world graph, which is inappropriate for a large social network since its degree distribution is not heavy-tailed. An extension of this work should consider scale-free networks.

Author Contributions: Writing and visualization, Y.W.; supervision, P.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kim, H.K.; Tandoc, E.C. Consequences of Online Misinformation on COVID-19: Two Potential Pathways and Disparity by EHealth Literacy. *Front. Psychol.* **2022**, *13*, 783909. [[CrossRef](#)] [[PubMed](#)]
2. Barua, Z.; Barua, S.; Aktar, S.; Kabir, N.; Li, M. Effects of Misinformation on COVID-19 Individual Responses and Recommendations for Resilience of Disastrous Consequences of Misinformation. *Prog. Disaster Sci.* **2020**, *8*, 100119. [[CrossRef](#)] [[PubMed](#)]
3. Treen, K.M.D.; Williams, H.T.P.; O'Neill, S.J. Online Misinformation about Climate Change. *WIREs Clim. Chang.* **2020**, *11*, e665. [[CrossRef](#)]
4. Acemoglu, D.; Ozdaglar, A. Opinion Dynamics and Learning in Social Networks. *Dyn. Games Appl.* **2010**, *1*, 3–49. [[CrossRef](#)]
5. Mostagir, M.; Siderius, J. Learning in a Post-Truth World. *Manag. Sci.* **2022**, *68*, 4340. [[CrossRef](#)]
6. Nguyen, N.P.; Yan, G.; Thai, M.T. Analysis of misinformation containment in online social networks. *Comput. Netw.* **2013**, *57*, 2133–2146. [[CrossRef](#)]
7. Zareie, A.; Sakellariou, R. Minimizing the Spread of Misinformation in Online Social Networks: A Survey. *J. Netw. Comput. Appl.* **2021**, *186*, 103094. [[CrossRef](#)]
8. Mostagir, M.; Siderius, J. Social Inequality and the Spread of Misinformation. *Manag. Sci.* **2022**, *69*, 968–995. [[CrossRef](#)]
9. Martins, A.C.R. Continuous Opinions and Discrete Actions in Opinion Dynamics Problem. *Int. J. Mod. Phys. C* **2008**, *19*, 617–624. [[CrossRef](#)]

10. Martins, A.C.R. Discrete Opinion Models as a Limit Case of the CODA Model. *Phys. A Stat. Mech. Appl.* **2014**, *395*, 352–357. [[CrossRef](#)]
11. Martins, A.C.R. Discrete Opinion Dynamics with M Choices. *Eur. Phys. J. B* **2020**, *93*, 1–10. [[CrossRef](#)]
12. Liu, M.; Rong, L. An Online Multi-Dimensional Opinion Dynamic Model with Misinformation Diffusion in Emergency Events. *J. Inf. Sci.* **2021**, *48*, 640–659. [[CrossRef](#)]
13. Koprulu, I.; Kim, Y.; Shroff, N.B. Battle of Opinions over Evolving Social Networks. *IEEE/ACM Trans. Netw.* **2019**, *27*, 532–545. [[CrossRef](#)]
14. Hegselmann, R.; Krause, U. Opinion Dynamics under the Influence of Radical Groups, Charismatic Leaders, and Other Constant Signals: A Simple Unifying Model. *Netw. Heterog. Media* **2015**, *10*, 477–509. [[CrossRef](#)]
15. Zhang, J.; Chen, G. Convergence Rate of the Asymmetric Deffuant-Weisbuch Dynamics. *J. Syst. Sci. Complex.* **2015**, *28*, 773–787. [[CrossRef](#)]
16. Douven, I. Optimizing Group Learning: An Evolutionary Computing Approach. *Artif. Intell.* **2019**, *275*, 235–251. [[CrossRef](#)]
17. Altafini, C.; Ceragioli, F. Signed Bounded Confidence Models for Opinion Dynamics. *Automatica* **2018**, *93*, 114–125. [[CrossRef](#)]
18. Douven, I.; Hegselmann, R. Mis- and Disinformation in a Bounded Confidence Model. *Artif. Intell.* **2020**, *291*, 103415. [[CrossRef](#)]
19. Krause, U.; Stöckler, M. (Eds.) *Modellierung und Simulation von Dynamiken mit Vielen Interagierenden Akteuren*; Modus; University of Bremen: Bremen, Germany, 1997.
20. Krause, U. A Discrete Nonlinear and Non-autonomous Model of Consensus Formation. *Commun. Differ. Equ.* **2000**, *2000*, 227–236.
21. Hegselmann, R.; Krause, U. Opinion Dynamics and Bounded Confidence Models, Analysis and Simulation. *J. Artif. Soc. Soc. Simul.* **2002**, *5*.
22. Deffuant, G.; Neau, D.; Amblard, F.; Weisbuch, G. Mixing Beliefs among Interacting Agents. *Adv. Complex Syst.* **2000**, *3*, 87–98. [[CrossRef](#)]
23. Weisbuch, G.; Deffuant, G.; Amblard, F.; Nadal, J.-P. Meet, Discuss, and Segregate! *Complexity* **2002**, *7*, 55–63. [[CrossRef](#)]
24. Fortunato, S. Universality of the Threshold for Complete Consensus for the Opinion Dynamics of Deffuant et Al. *Int. J. Mod. Phys. C* **2004**, *15*, 1301–1307. [[CrossRef](#)]
25. Su, W.; Chen, G.; Ye, H. Noise Leads to Quasi-Consensus of Hegselmann–Krause Opinion Dynamics. *Automatica* **2017**, *85*, 448–454. [[CrossRef](#)]
26. Huang, C.; Dai, Q.; Han, W.; Feng, Y.; Cheng, H.; Li, H. Effects of Heterogeneous Convergence Rate on Consensus in Opinion Dynamics. *Phys. A Stat. Mech. Its Appl.* **2018**, *499*, 428–435. [[CrossRef](#)]
27. Lorenz, J. Heterogeneous Bounds of Confidence: Meet, Discuss and Find Consensus! *Complexity* **2009**, *15*, 43–52. [[CrossRef](#)]
28. Chazelle, B.; Wang, C. Inertial Hegselmann-Krause Systems. *IEEE Trans. Autom. Control.* **2017**, *62*, 3905–3913. [[CrossRef](#)]
29. Lorenz, J. Continuous Opinion Dynamics under Bounded Confidence: A Survey. *Int. J. Mod. Phys. C* **2007**, *18*, 1819–1838. [[CrossRef](#)]
30. Mirtabatabaei, A.; Bullo, F. Opinion Dynamics in Heterogeneous Networks: Convergence Conjectures and Theorems. *SIAM J. Control. Optim.* **2012**, *50*, 2763–2785. [[CrossRef](#)]
31. Banisch, S.; Lima, R.; Araújo, T. Agent Based Models and Opinion Dynamics as Markov Chains. *Soc. Netw.* **2012**, *34*, 549–561. [[CrossRef](#)]
32. Shang, Y. An Agent Based Model for Opinion Dynamics with Random Confidence Threshold. *Commun. Nonlinear Sci. Numer. Simul.* **2014**, *19*, 3766–3777. [[CrossRef](#)]
33. Pineda, M.; Buendía, G.M. Mass Media and Heterogeneous Bounds of Confidence in Continuous Opinion Dynamics. *Phys. A Stat. Mech. Its Appl.* **2015**, *420*, 73–84. [[CrossRef](#)]
34. Chen, G.; Su, W.; Mei, W.; Bullo, F. Convergence Properties of the Heterogeneous Deffuant–Weisbuch Model. *Automatica* **2020**, *114*, 108825. [[CrossRef](#)]
35. Sobkowicz, P. Extremism without Extremists: Deffuant Model with Emotions. *Front. Phys.* **2015**, *3*, 17. [[CrossRef](#)]
36. Li, S.; Zhang, S. The Leader and Extremist in the Bounded Confidence Model on Flocking Movement World. In Proceedings of the 2011 IEEE International Conference on Computer Science and Automation Engineering (CSAE), Shanghai, China, 10–12 June 2011.
37. Chen, S.; Glass, D.H.; McCartney, M. Characteristics of Successful Opinion Leaders in a Bounded Confidence Model. *Phys. A Stat. Mech. Its Appl.* **2016**, *449*, 426–436. [[CrossRef](#)]
38. Brooks, H.Z.; Porter, M.A. A Model for the Influence of Media on the Ideology of Content in Online Social Networks. *Phys. Rev. Res.* **2020**, *2*, 023041. [[CrossRef](#)]
39. Huang, S.; Xiu, B.; Feng, Y. Modeling and Simulation Research on Propagation of Public Opinion. In Proceedings of the 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, China, 3–5 October 2016.
40. Tian, Y.; Wang, L. Opinion Dynamics in Social Networks with Stubborn Agents: An Issue-Based Perspective. *Automatica* **2018**, *96*, 213–223. [[CrossRef](#)]
41. Knop, K.; Öncü, J.S.; Penzel, J.; Abele, T.S.; Brunner, T.; Vorderer, P.; Wessler, H. Offline Time is Quality Time. Comparing Within-Group Self-Disclosure in Mobile Messaging Applications and Face-To-Face Interactions. *Comput. Hum. Behav.* **2016**, *55*, 1076–1084. [[CrossRef](#)]

42. Golub, B.; Jackson, M.O. Naïve Learning in Social Networks and the Wisdom of Crowds. *Am. Econ. J. Microecon.* **2010**, *2*, 112–149. [[CrossRef](#)]
43. Banerjee, A.; Fudenberg, D. Word-of-Mouth Learning. *Games Econ. Behav.* **2004**, *46*, 1–22. [[CrossRef](#)]
44. Acemoglu, D.; Dahleh, M.A.; Lobel, I.; Ozdaglar, A. Bayesian Learning in Social Networks. *Rev. Econ. Stud.* **2011**, *78*, 1201–1236. [[CrossRef](#)]
45. Pineda, M.; Toral, R.; Hernández-García, E. The Noisy Hegselmann-Krause Model for Opinion Dynamics. *Eur. Phys. J. B* **2013**, *86*, 490. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.