# Report: Modeling Misinformation Diffusion in NetLogo, The Role of Agent Susceptibility and Fact-Checking Behavior

Erwin Terpstra, Ning Wei Zhou, Jorrit Adema
Open University of the Netherlands

Abstract

The large-scale dissemination of disinformation on social media networks is a real threat to public discussion and decision-making. We used an agent-based simulation in NetLogo to examine the effects of susceptibility to disinformation and fact-checking on the spread and duration of false rumors. Based on an existing SIS-based model, we introduced heterogeneous agent susceptibility, a fact-checking threshold, and anchoring bias, modeled as a reduced forgetting rate. The experiments contrast the baseline model with versions incorporating these variables. Results show higher susceptibility significantly increases long-term prevalence of misinformation. Anchoring bias enhances the effect when agents generally do not fact-check. Heterogeneous susceptibility slows early spread but allows misinformation to persist longer in isolated clusters. Fact-checking lowers long-term belief in untruth but can increase short-term spread. Anchoring bias has context-dependent effects, widening or narrowing spread as a function of verification levels. These findings highlight the requirement for personal cognitive traits to shape misinformation dynamics and suggest that countermeasures need to be responsive to varying levels of user susceptibility and trust in verification. Correspondingly, all experimental datasets and analysis scripts underlying this study are publicly accessible at `https://github.com/ErwinTerpstra/MisinformationSusceptibility`.

## Background

In the internet era, the rapid dissemination of information has fundamentally changed the way humans interact with news, opinions, and knowledge. Social networking websites such as Facebook, X (formerly Twitter), and Instagram have become critical channels for the exchange of everyday life experiences, thoughts, and opinions, with billions of users worldwide contributing to an ever-increasing information ecosystem (Pew Research Center, 2021).

But the same velocity and scale of digital communication have also facilitated the spread of knowledge of misinformation, which is false or inaccurate information communicated without intention to deceive. Misinformation includes rumors, hoaxes, and fake news and is likely to occur in the context of inadequate verification (Friggeri, Adamic, Eckles, & Cheng, 2014).

Vosoughi et al demonstrated in their study that false claims are 70% more likely to be retweeted than factual claims (Vosoughi, Roy, & Aral, 2018). That misinformation becomes viral is even more concerning since it has the ability to influence public opinion, disrupt democratic processes, and impact public health, as demonstrated by the viral spread of vaccine misinformation (J. Li & Chang, 2022). The dangerous impacts of misinformation are worsened by the fact that individuals do not know the fake content that they receive and share, leading to the dissemination of falsehoods (Pennycook & Rand, 2018).

In reaction to the increasing issue of misinformation, researchers have found a number of solutions, including fact-checking initiatives, media literacy campaigns, and algorithmic interventions to limit the spread of misleading content (Pennycook & Rand, 2018). Fact-checking, in particular, has been identified as one of the most promising interventions for misinformation, with studies showing that debunking misinformation can significantly reduce its impact (Micallef, He, Kumar, Ahamad, & Memon, 2020). However, despite these efforts, the effectiveness of interventions remains an issue of controversial debate, with some studies showing that correcting misinformation may inadvertently reinforce false beliefs due to the backfire effect (Nyhan & Reifler, 2010).

This phenomenon has spurred extensive research into its mechanisms and potential countermeasures. J. Li and Chang (2022) demonstrated that fact-checked statements framed as entirely true or entirely false tend to be shared more widely than those deemed partially true, indicating a preference for clear, unambiguous classifications. In addition, this study emphasizes that the credibility of the source is a critical factor. Less reliable sources, often due to their sensational content, capture more attention and that repeated dissemination through social media further amplifies exposure.

Building on these insights, Z. Li, Cao, Adams-Cohen, and Alvarez (2023) investigated how Donald Trump's tweets affect misinformation spread and evaluated the impact of moderation strategies. Their findings reveal that Trump's tweets typically trigger a sharp surge in misinformation, with volumes peaking soon after posting and stabilizing at higher levels, thereby directly contributing to the phenomenon. While Twitter's restrictions generally reduced misinformation, warning labels yielded more mixed outcomes; in some cases, such labels even coincided with a dramatic 1600% increase in misinformation dissemination, highlighting the challenges of implementing effective mitigation measures.

Other research has helped to explain processes through which misinformation spreads in social networks. For instance, Zollo et al. (2019) demonstrated how confirmation bias and echo chambers fuel misinformation diffusion in digital spaces by citing that polarized groups tend to be resistant to efforts against fact-checking. (Del Vicario, Scala, Caldarelli, Stanley, & Quattrociocchi, 2016) similarly pointed out that misinformation spreads in homogeneous social networks with like-minded individuals repeating unverified claims to one another, rendering corrective interventions less relevant. The studies emphasize network structure and user engagement patterns in influencing misinformation dissemination.

Motivated by these empirical findings, this study employs simulation-based experiments to further examine the dynamics of misinformation diffusion and to evaluate the effects of anchoring bias and susceptibility thresholds. By integrating simulation with empirical data, this research specifically investigates how introducing agent-specific susceptibility and anchoring bias affects misinformation diffusion and the effectiveness of fact-checking interventions in a simulated environment.

These simulations aim to test how agent-level susceptibility and anchoring bias influence misinformation dynamics, and to assess how these characteristics shape the overall impact of internal fact-checking behavior, enabling a deeper exploration of their implications and effectiveness in addressing the challenges posed by misinformation.

## Research questions

Previous studies have demonstrated that user behavior plays a critical role in the dissemination of misinformation, particularly through mechanisms such as selective exposure and responses to content verification (Sulis & Tambuscio, 2020; Tambuscio, Ruffo, Flammini, & Menczer, 2015). Nonetheless, less attention has been paid to the influence of individual-level characteristics, such as inherent susceptibility to misinformation and the propensity to engage in fact-checking. A survey conducted by Eurostat reported that only about 23% of Europeans actively verify the information they consume online (Eurostat, 2021).

By explicitly modeling heterogeneous susceptibility and its impact on verification behavior, this work aims to investigate the spread of misinformation within a population of diverse individual responses and evaluate the effectiveness of fact-checking behavior under varying susceptibility and anchoring conditions.

Moreover, previous research has shown that the psychosocial phenomenon known as 'anchoring bias' is strongly associated with the propagation of misinformation (Jost, Pünder, & Schulze-Lohoff, 2020; Sherif, Taub, & Hovland, 1958). Anchoring bias refers to the tendency of individuals to form initial beliefs based on the first piece of information encountered, which in the context of misinformation, may lead to an increased likelihood of accepting hoaxes among agents who have not yet solidified their opinions on the subject. In view of these considerations, the research will address the following questions:

**RQ1.** How does the introduction of agent-specific susceptibility influence the persistence of misinformation?

**RQ2.** To what extent does heterogeneous susceptibility affect the spreading rate of misinformation?

**RQ3.** To what extent do agents above the verification threshold (i.e., who never verify) influence the effectiveness of fact-checking interventions?

**RQ4.** What is the effect of incorporating anchoring bias on the spread of misinformation?

## Research methods

### Simulation

The research is based on a multi-agent simulation using the NetLogo software (version 6.4.0). The simulation model is of adapted form of the model developed by Sulis and Tambuscio (2020). The enhanced model will simulate differences between individuals in their susceptibility to viral information, as well as their natural propensity to verify information with a trusted source.

Sulis' model is built as an analogy to a SIS model, which is widely employed in simulations to study the dynamics of infectious disease propagation. The SIS (Susceptible-Infected-Susceptible) model, widely used in epidemiology, provides a useful analogy for misinformation dynamics, where individuals may adopt misinformation, forget it, and become susceptible again.

Within this model, agents exist in one of three distinct states:

- **Susceptible (S)**: Agents who have not adopted any belief yet.
- **Believer (B)**: Agents who are spreading misinformation.
- **FactChecker (FC)**: Agents who reject misinformation. This does not imply professional fact-checkers.

During each simulation interval, defined by $t \in \{1, 2, 3, ..., 300\}$, agents may either persist in their current state or transition to an alternative state. The model delineates three distinct phenomena governing state transitions.

- **Spreading** $[S \to B,\ S \to FC]$ Each agent is characterized by a probability of transitioning from the S state to either the B state or the FC state.
- **Verifying** $[B \to FC]$ Each agent is assigned a fixed probability, $p_{verify}$, to assess the accuracy of misinformation and, subsequently, alter its state to FC state.
- **Forgetting** $[B \to S,\ FC \to S]$ Each agent possesses a fixed probability, $p_{forget}$, to discard information and return to the S state.

The probabilities of state transitions from the S state are given by the following expressions for going from S to B (spreading) or from S to FC (debunking):

$$f_i(t) = \beta \frac{n_i^B(t)(1+\alpha)}{n_i^B(t)(1+\alpha) + n_i^{FC}(t)(1-\alpha)} \quad \text{(spreading)}$$
$$g_i(t) = \beta \frac{n_i^{FC}(t)(1-\alpha)}{n_i^B(t)(1+\alpha) + n_i^{FC}(t)(1-\alpha)} \quad \text{(debunking)}$$

*Note:* In this model, we refer to the transition from $B \to FC$ as "verifying" because the agent revises a previously held belief. The transition from $S \to FC$ is called "debunking" as it represents the agent rejecting misinformation at first exposure. Both transitions reflect fact-checking behavior, but differ in timing relative to belief adoption.

Remaining in the same state at time $t$: These equations represent the probabilities that agent $i$ remains in its current state.

$$\begin{aligned} P_{S,i}(t) &= 1 - f_i(t) - g_i(t), \\ P_{B,i}(t) &= 1 - p_{forget} - p_{verify}, \\ P_{FC,i}(t) &= 1 - p_{forget}. \end{aligned}$$

In the context of our model, the parameter $\beta \in [0,1]$ denotes the spread rate, which quantifies the likelihood of information dissemination throughout the network. The parameter $\alpha \in [0,1]$, on the other hand, measures the credibility of the hoax under evaluation. Additionally, $n_i^B$ and $n_i^{FC}$ represent the number of neighboring agents classified as Believer and FactChecker, respectively, at time $t$.

Notably, the model is constrained such that $f_i(t) + g_i(t) = \beta$ holds at any given time $t$, thereby maintaining consistency in the overall diffusion dynamics. In summary, four key parameters form the basis of this model: the spreading rate $\beta$, the hoax credibility $\alpha$, the probability $p_{verify}$ of fact-checking a hoax, and the probability $p_{forget}$ of an agent forgetting its current belief. These elements collectively define the structure of the model as depicted in Figure 1.

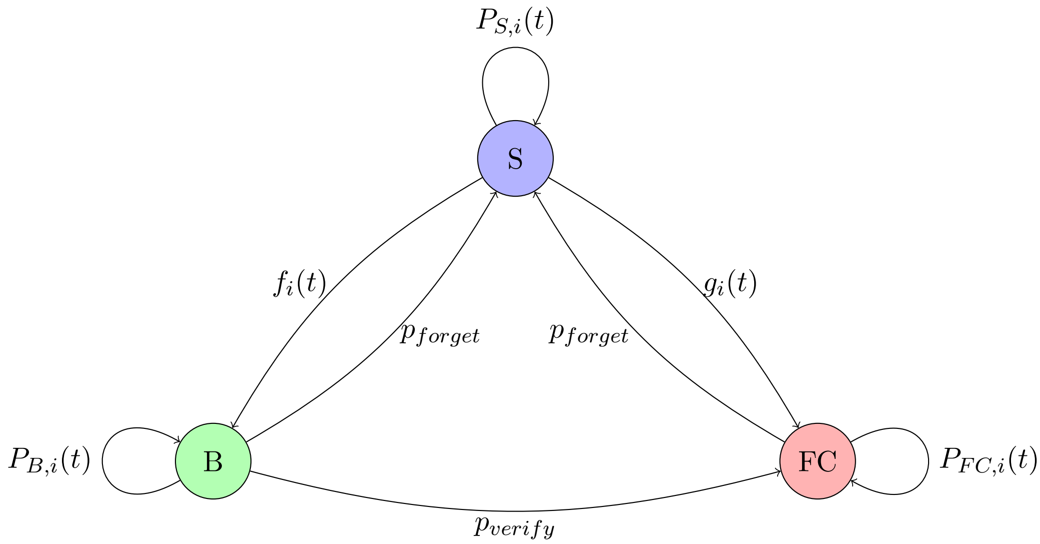



*Figure 1.* States and transitions of the original model

## Enhanced model

In the original model, it is assumed that individual behavior is uniformly distributed across all agents, such that each agent demonstrates identical levels of responsiveness to external stimuli. However, empirical evidence, by Shiri, Hatef, and Sahraei (2016), suggests that individuals often differ significantly in their propensity to change states. To capture this heterogeneity, we extend the original model by introducing two variables: an agent-specific susceptibility factor, denoted by $s_f \in [0,1]$, and a model-wide anchoring factor, denoted by $a_f \in \{0,1\}$.

The susceptibility factor $s_f$ quantifies an agent's inherent tendency to adopt a new state in response to external influences, whereas the anchoring factor $a_f$ reflects whether an agent's susceptibility factor also influences their tendency to retain their current state, despite environmental perturbations. To integrate these variable influences into the model, several key parameters are modulated by scaling with the susceptibility factor. The enhanced parameters are defined by:

$$
\begin{aligned}
f'_i(t) &= f_i(t) \cdot s_f \\
g'_i(t) &= g_i(t) \cdot (1 - s_f) \\
p'_{verify} &= \begin{cases} p_{verify} \cdot (1 - s_f), & \text{if } s_f \leq s_{max-verify} \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
$$

Where $s_{max-verify}$ denote the susceptibility threshold above which an agent will never perform information verification. Since the susceptibility $s_f$ is uniformly distributed, $s_{max-verify}$ effectively represents the fraction of the population that is willing to verify information—that is, it corresponds to the maximum susceptibility observed within the population. In Experiments EX2 and EX3, a random susceptibility value is assigned for all agents in each run.

Additionally, Anchoring Bias refers to the phenomenon whereby agents assign disproportionate importance to the initial information they encounter, often accepting it as truth. In our model, when the anchoring mechanism is activated (i.e., $a_f = 1$), this bias is incorporated by modifying the baseline forgetting probability $p_{\text{forget}}$ using a susceptibility factor, $(1 - s_f)$. Consequently, the adjusted forgetting probability, denoted as $p'_{\text{forget}}$, is defined as follows:

$$
p'_{forget} = \begin{cases} p_{forget} \cdot (1 - s_f), & \text{if } a_f = 1 \\ p_{forget} & \text{otherwise} \end{cases}
$$

Consequently, the probabilities of an agent remaining in each current state are updated as follows:

$$
\begin{aligned}
P'_{S,i}(t) &= 1 - f'_i(t) - g'_i(t), \\
P'_{B,i}(t) &= 1 - p'_{forget} - p'_{verify}, \\
P'_{FC,i}(t) &= 1 - p'_{forget}.
\end{aligned}
$$

These modifications allow the model to account for individual variability in susceptibility to external influences as well as the differential stability of agents' states under changing conditions.

## Experimental Framework

In this section, we detail the experiments designed to evaluate our proposed model under controlled conditions and to determine whether the simulation outcomes validate our hypotheses.

The simulation is executed for a total of 300 ticks (i.e., $t_{max} = 300$) to match with the experiments performed by Sulis and Tambuscio (2020). During this period, the number of agents in each of the three states is continuously monitored. Moreover, each experiment is repeated for 300 independent runs to mitigate the effects of stochastic variability. This number has been chosen by manual experimentation to be the level at which consistent results could be obtained.

In the initial experiment, we implement the original model as described by Sulis and Tambuscio (2020), utilizing the parameter values defined therein to establish a robust baseline. Subsequently, we extend our investigation by deploying an enhanced version of the model that incorporates the $s_{max-verify}$ variable, thereby allowing us to examine its influence on simulation outcomes.

Finally, we conduct additional experiment in which the enhanced model is executed with anchoring bias enabled, facilitating an evaluation of its impact on the dynamics of misinformation spread. For clarity, the experimental procedures are delineated as follows:

1. **Experiment 1 (EX1, Baseline):** Run the original model without incorporating any of the proposed enhancements.
2. **Experiment 2 (EX2):** Execute the enhanced model with varying values of $s_{max-verify}$.
3. **Experiment 3 (EX3):** Execute the enhanced model with varying values of $s_{max-verify}$ while enabling the anchoring bias extension.

**Procedure**

The experiments are conducted utilizing the NetLogo BehaviorSpace tool, which allows for the pre-configuration of experimental variables followed by a defined number of repetitions for each configuration. Data from every individual run is automatically recorded, and upon the culmination of each simulation, the results are exported to a CSV file. This approach enables both the analysis of the final simulation state and an examination of the simulation's temporal dynamics.

**Parameter Settings**

To assess the influence of the verification threshold $s_{max\text{-}verify}$ on model outcomes, we conducted ten simulation runs for experiments EX2 and EX3, varying $s_{max\text{-}verify}$ systematically from 0.1 to 1.0 in increments of 0.1. Table 1 summarizes these values. All other model parameters remained fixed.

Table 1
*Values for $s_{max\text{-}verify}$ used in experiments EX2 and EX3.*

| **Run (R)** | $s_{max\text{-}verify}$ |
|---|---|
| 1 | 0.1 |
| 2 | 0.2 |
| 3 | 0.3 |
| 4 | 0.4 |
| 5 | 0.5 |
| 6 | 0.6 |
| 7 | 0.7 |
| 8 | 0.8 |
| 9 | 0.9 |
| 10 | 1.0 |

The remaining parameters follow the baseline configuration established by Sulis et al. Table 2 lists each constant parameter, its assigned value, and a concise explanation of its role in the simulation.

Table 2
*Baseline constant parameter settings.*

| **Variable** | **Value** | **Explanation** |
|---|---|---|
| number-of-agents ($N_{\text{agents}}$) | 1000 | Total number of agents in the simulation. |
| min-links-per-agent | 3 | Minimum number of links (connectivity) per agent in the network. |
| type-of-network | Barabási-Albert | Network topology generated using the Barabási-Albert algorithm. |
| $p_{\text{setupB}}$ | 0.10 | Proportion of agents initialized as 'Believers' at the start of the simulation. |
| $p_{\text{forget}}$ | 0.10 | Probability that an agent in the 'Believer' or 'Fact-checker' state reverts to the neutral 'Susceptible' state. |
| $p_{\text{verify}}$ | 0.05 | Baseline probability that a 'Believer' fact-checks the information, transitioning to the 'Fact-checker' state. |
| $\beta$ | 0.50 | Base probability that a 'Susceptible' agent adopts information from its neighbors. |
| $\alpha$ | 0.25 | Relative likelihood of accepting a hoax over fact-checked information, making a hoax approximately 67% more credible. |

**Data Analysis**

The simulation output will be processed using Python, with key functionalities provided by packages such as Matplotlib, Pandas, and SciPy. Emphasis will be placed on robust data visualization techniques and comprehensive statistical comparisons in order to establish the statistical significance of any observed effects. To further facilitate the interpretation of our results, we introduce two derived variables: $hoax_{pe}$ and $hoax_{sp}$.

1. **Hoax Persistence**: This metric quantifies the relative proportion of agents endorsing the hoax at the simulation's conclusion. For each experimental repetition, the persistence is determined by computing the average fraction of "believer" agents over the final 10 time ticks of the simulation. Formally, this is expressed as:

$$hoax_{\text{pe}} = \frac{1}{10} \sum_{t=t_{\max}-9}^{t_{\max}} \frac{B_t}{N_{\text{agents}}},$$

where $B_t$ represents the number of agents in the "believer" state at time $t$ and $N_{\text{agents}}$ denotes the total number of agents in the simulation. The unit of this metric is a fraction of believers compared to the total population.

2. **Hoax Spreading Rate**: This metric measures the initial velocity of hoax diffusion. It is calculated by determining the time at which the number of believer agents reaches its maximum and then computing the rate of increase from the initial state. This rate is defined as:

$$hoax_{\text{sp}} = \frac{B_{\max} - B_{\text{initial}}}{t_{\max}},$$

where $B_{\max}$ is the maximum number of agents in the "believer" state recorded during the simulation, and $B_{\text{initial}}$ is the initial count of agents in this state. The unit of this metric is agents per simulation tick.

In this experimental framework, the variable $B_t$ denotes the number of agents in the "believer" state at time $t$. The metric $B_{\max}$ represents the peak number of believer agents observed over the entire simulation period, while $B_{\text{initial}}$ indicates the initial count of believer agents at the commencement of the simulation. These derived metrics provide critical insights for identifying and quantifying variations in simulation behavior under different experimental configurations.

## Hypothesis

Based on our experimental framework, theoretical model, and the previously articulated research questions, we propose a set of interrelated hypotheses to elucidate the influence of individual cognitive characteristics on the trajectories of misinformation propagation. These conjectures have been operationalized into specific null and alternative hypotheses. For clarity, each hypothesis is presented with its formal testing framework. In addition, t-tests are applied to each testing framework to assess the statistical significance of the observed outcomes.

### A. Persistence (Retention of Misleading Beliefs)

Persistence is measured by the metric $hoax_{pe}$. We consider two main conditions:

**Hypothesis 1a: Susceptibility Only:** We test whether inherent differences in individual susceptibility alone lead to changes in the persistence of misleading beliefs.

$$H_0: \ hoax_{pe_{ex1}} = hoax_{pe_{ex2}}, \quad H_1: \ hoax_{pe_{ex1}} \neq hoax_{pe_{ex2}}$$

**Hypothesis 1b: With Anchoring Bias:** We determine if the introduction of an anchoring bias further affects the retention of misinformation compared to the baseline condition.

$$H_0: \ hoax_{pe_{ex1}} = hoax_{pe_{ex3}}, \quad H_1: \ hoax_{pe_{ex1}} \neq hoax_{pe_{ex3}}$$

### B. Spread (Diffusion of Misinformation)

Spread is measured by the metric $hoax_{sp}$. Similarly, we examine two conditions:

**Hypothesis 2a: Heterogeneity Only:** We explore the effect of individual heterogeneity in susceptibility on the rate of misinformation diffusion.

$$H_0: \ hoax_{sp_{ex1}} = hoax_{sp_{ex2}}, \quad H_1: \ hoax_{sp_{ex1}} \neq hoax_{sp_{ex2}}$$

**Hypothesis 2b: With Anchoring Bias:** We evaluate whether the introduction of an anchoring bias in heterogeneous populations further accelerates or modifies the diffusion process.

$$H_0 : \ hoax_{sp_{ex1}} = hoax_{sp_{ex3}}, \quad H_1 : \ hoax_{sp_{ex1}} \neq hoax_{sp_{ex3}}$$

**C. Consistency Across Experimental Runs (Experiment 2)**

For Experiment 2, we assess the consistency of both persistence and spread metrics across multiple runs ($R = 1, \ldots, 10$). These runs correspond to different parameter settings as specified in Table 1, which detail variations in the underlying agent characteristics.

**Hypothesis 3a: Persistence Consistency:** We verify that the persistence measure $hoax_{pe_{ex2}}$ remains consistent over the different parameter settings.

$$\begin{aligned} H_0 : &\quad \forall\, R_x, R_y \in \{1, ..., 10\}, R_x \neq R_y, \ hoax_{pe_{ex2}}(R_x) = hoax_{pe_{ex2}}(R_y) \\ H_1 : &\quad \exists\, R_x, R_y \in \{1, ..., 10\}, R_x \neq R_y \text{ such that } hoax_{pe_{ex2}}(R_x) \neq hoax_{pe_{ex2}}(R_y) \end{aligned}$$

**Hypothesis 3b: Spread Consistency:** We test the invariance of the diffusion measure $hoax_{sp_{ex2}}$ across the different parameter settings.

$$\begin{aligned} H_0 : &\quad \forall\, R_x, R_y \in \{1, ..., 10\}, R_x \neq R_y, \ hoax_{sp_{ex2}}(R_x) = hoax_{sp_{ex2}}(R_y) \\ H_1 : &\quad \exists\, R_x, R_y \in \{1, ..., 10\}, R_x \neq R_y \text{ such that } hoax_{sp_{ex2}}(R_x) \neq hoax_{sp_{ex2}}(R_y) \end{aligned}$$

**D. Relative Impact of Anchoring Bias (Experiment 2 vs. Experiment 3)**

Finally, we compare the experiments conducted with and without the anchoring bias to evaluate its differential effects. The experimental runs ($R = 1, \ldots, 10$) correspond to distinct conditions characterized by varying parameter settings, as detailed in Table 1.

**Hypothesis 4a: Relative Impact on Persistence:** We compare $hoax_{pe}$ between Experiment 2 (without anchoring bias) and Experiment 3 (with anchoring bias). The null hypothesis assumes that there exists at least one run where the relative ordering of $hoax_{pe}$ is not uniform. In contrast, the alternative hypothesis maintains that the relative ordering of $hoax_{pe}$ is consistent across all runs.

$$\begin{aligned} H_0 : &\quad \exists\, R_x, R_y \in \{1, ..., 10\} \text{ such that } hoax_{pe_{ex2}}(R_x) \geq hoax_{pe_{ex3}}(R_y) \\ H_1 : &\quad \forall\, R_x, R_y \in \{1, ..., 10\} \text{ we have that } hoax_{pe_{ex2}}(R_x) < hoax_{pe_{ex3}}(R_y) \end{aligned}$$

**Hypothesis 4b: Relative Impact on Spread:** We similarly compare the diffusion measures $hoax_{sp}$ between Experiment 2 and Experiment 3.

$$\begin{aligned} H_0 : &\quad \exists\, R_x, R_y \in \{1, ..., 10\} \text{ such that } hoax_{sp_{ex2}}(R_x) \geq hoax_{sp_{ex3}}(R_y) \\ H_1 : &\quad \forall\, R_x, R_y \in \{1, ..., 10\} \text{ we have that } hoax_{sp_{ex2}}(R_x) < hoax_{sp_{ex3}}(R_y) \end{aligned}$$

**Student's $t$–test**

Having delineated our hypotheses regarding persistence, spread, consistency across experimental runs, and the relative impact of the anchoring bias, we proceed to compute the relevant test statistics. We employ Student's $t$–test to determine whether observed differences in sample means could plausibly arise under the null hypothesis of no true effect.

**Tail specification.** Hypotheses 1a, 1b, 2a, 2b, 3a and 3b admit deviations in either direction and are therefore evaluated with two-tailed $t$–tests. Hypothesis 4a and 4b posits a directional effect and is evaluated with a one-tailed $t$–test.

**One-sample $t$–test.**

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}},$$

where $\overline{x}$ is the sample mean, $\mu_0$ the hypothesized population mean, $s$ the sample standard deviation, and $n$ the sample size.

**Two-sample $t$–test (equal variances).**

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad s_p = \sqrt{\frac{(n_1 - 1)\, s_1^2 \; + \; (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}},$$

where $\overline{x}_i$, $s_i$, and $n_i$ denote the sample mean, standard deviation, and size of group $i$, and $s_p$ is the pooled standard deviation.

In the next section, we will present a systematic analysis of the experimental data, compare key metrics across the various conditions, and evaluate the degree to which the empirical evidence supports the stated hypotheses. The insights gained from this analysis will form the basis for the conclusions addressing our research questions.

## Results

All experimental datasets and analysis scripts underlying this study are publicly accessible at `https://github.com/ErwinTerpstra/MisinformationSusceptibility`.

To evaluate hypotheses 1a, 1b, 2a, and 2b, we examined the derived variables $hoax_{pe}$ (persistence) and $hoax_{sp}$ (spread) at the reference verification threshold $s_{\text{max-verify}} = 0.2$. Figure 2 depicts the distributions of these metrics across all experimental runs, with means and variances computed over repeated trials. Pairwise two-sided $t$-tests demonstrate that both EX2 (heterogeneous susceptibility) and EX3 (anchoring bias) differ significantly from the uniform-susceptibility baseline EX1 in terms of persistence and spread. The resulting $p$-values, each indicating the probability that the two compared samples derive from the same underlying distribution—are reported in Table 3.

Table 3

*P-values for hoax persistence ($hoax_{pe}$) and spread ($hoax_{sp}$) at the reference threshold $s_{\max\text{-verify}} = 0.2$, based on two-sided t-tests.*

| **Experiment pair** | **P-value** $hoax_{pe}$ | **P-value** $hoax_{sp}$ |
|---|---|---|
| EX1 vs. EX2 | $p < 0.001$ | $p < 0.001$ |
| EX1 vs. EX3 | $p < 0.001$ | $p < 0.001$ |
| EX2 vs. EX3 | $p < 0.001$ | $p \approx 0.003$ |

To evaluate hypotheses 3a, 3b, 4a, and 4b, we compared the derived metrics $hoax_{pe}$ (persistence) and $hoax_{sp}$ (spread) between EX2 and EX3 over the entire range of $s_{\max\text{-verify}}$ values. Figure 3 shows these metrics with error bars representing variance across repeated trials. Statistical significance was assessed by performing pairwise, two-sided $t$-tests for every combination of runs, producing a symmetric matrix of $p$-values. Tables 4 and 5 present the lower-triangular portion of this matrix, omitting diagonal self-comparisons and redundant upper-triangle entries due to the commutativity of the $t$-test.

Table 4

*Pairwise two-sided t-tests for hoax persistence samples within EX2 runs. Values lower than 0.001 are rounded down to zero for readability.*

| **Run (R)** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | – | | | | | | | | | |
| 2 | 0 | – | | | | | | | | |
| 3 | 0 | 0 | – | | | | | | | |
| 4 | 0 | 0 | 0 | – | | | | | | |
| 5 | 0 | 0 | 0 | 0 | – | | | | | |
| 6 | 0 | 0 | 0 | 0 | 0 | – | | | | |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | – | | | |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | – | | |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | – | |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.056 | – |

Table 5

*Pairwise two-sided t-tests for hoax spread samples within EX2 runs. Values lower than 0.001 are rounded down to zero for readability.*

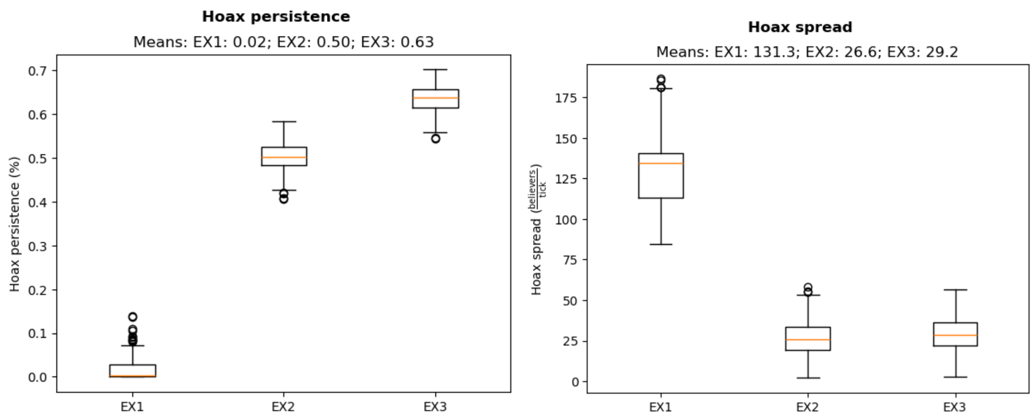
| **Run (R)** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | – | | | | | | | | | |
| 2 | 0 | – | | | | | | | | |
| 3 | 0 | 0 | – | | | | | | | |
| 4 | 0 | 0 | 0 | – | | | | | | |
| 5 | 0 | 0 | 0 | 0 | – | | | | | |
| 6 | 0 | 0 | 0 | 0 | 0.018 | – | | | | |
| 7 | 0 | 0 | 0 | 0 | 0 | 0.330 | – | | | |
| 8 | 0 | 0 | 0 | 0 | 0 | 0.003 | 0.051 | – | | |
| 9 | 0 | 0 | 0 | 0 | 0 | 0.032 | 0.236 | 0.476 | – | |
| 10 | 0 | 0 | 0 | 0 | 0 | 0.036 | 0.280 | 0.358 | 0.862 | – |

In addition, EX2 and EX3 were compared on a per-run basis, yielding one-tailed $t$-tests in line with the directional expectations of hypotheses 4a and 4b. Table 6 reports the resulting significance levels for each run (and hence for each value of $s_{\text{max-verify}}$). These per-run comparisons reveal that the anchoring bias implemented in EX3 produces opposing effects on hoax persistence and spread at low versus high verification thresholds. This pattern suggests a non-linear interaction between verification behavior and anchoring bias.

Table 6

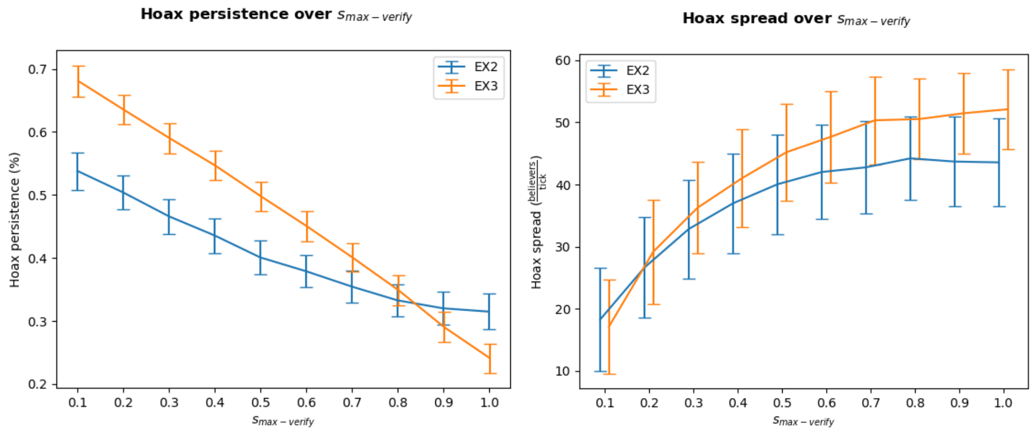
*One-tailed $t$-test results for $EX2 < EX3$ across different verification thresholds.*

| **Run (R)** | $s_{\text{max-verify}}$ | **P-value** $hoax_{pe}$ | **P-value** $hoax_{sp}$ |
|---|---|---|---|
| 1 | 0.1 | $p < 0.001$ | $p \approx 0.919$ |
| 2 | 0.2 | $p < 0.001$ | $p \approx 0.001$ |
| 3 | 0.3 | $p < 0.001$ | $p < 0.001$ |
| 4 | 0.4 | $p < 0.001$ | $p < 0.001$ |
| 5 | 0.5 | $p < 0.001$ | $p < 0.001$ |
| 6 | 0.6 | $p < 0.001$ | $p < 0.001$ |
| 7 | 0.7 | $p < 0.001$ | $p < 0.001$ |
| 8 | 0.8 | $p < 0.001$ | $p < 0.001$ |
| 9 | 0.9 | $p = 1.0$ | $p < 0.001$ |
| 10 | 1.0 | $p = 1.0$ | $p < 0.001$ |




(a) $hoax_{pe}$ (b) $hoax_{sp}$

*Figure 2*. Simulation metrics comparing EX2 and EX3 with baseline (EX1) at $s_{\text{max-}verify} = 0.2$.




(a) $hoax_{pe}$ (b) $hoax_{sp}$

*Figure 3*. Evolution of $hoax_{pe}$ and $hoax_{sp}$ for EX2 and EX3 over varying $s_{\max\text{-}verify}$.

## Conclusion

We examined how susceptibility, verification thresholds, and anchoring bias affect the spread and persistence of misinformation in agent-based simulations of social networks.

First, heterogeneous susceptibility significantly increased misinformation persistence. When 80% of agents never verified information (EX2), the proportion of persistent hoax believers ($hoax_{pe}$) was significantly higher than in a homogeneous setting (EX1, $p < 0.001$), confirming Hypotheses 1a and 1b. These results align with real-world observations: individuals with lower cognitive reflection or limited media literacy are more prone to believe and retain false information, contributing to long-lasting belief pockets, as seen in conspiracy theory communities or vaccine misinformation clusters.

Second, susceptibility heterogeneity reduced the initial spread of misinformation. The lower average adoption probability in the heterogeneous model led to a significantly smaller early spread ($hoax_{sp}$, $p < 0.001$), supporting Hypotheses 2a and 2b. This reflects how, in practice, misinformation does not always go viral instantly but may gradually embed itself in specific demographic or ideological groups over time.

Third, verification behavior had dual effects. Higher prevalence of verification reduced hoax persistence (confirming Hypothesis 3a) but also increased early spread ($hoax_{sp}$), consistent with Hypothesis 3b. This effect arises because verifying agents delay adoption, while misinformation circulates rapidly among non-verifiers. A real-world analogue is the rapid initial sharing of viral falsehoods on social media, often followed by slower fact-checking corrections, which may eventually reduce belief but not prevent initial exposure.

Fourth, anchoring bias produced context-dependent effects. In low-verification settings, anchoring increased persistence; in high-verification settings, it reduced it. Similarly, hoax spread rose more sharply under anchoring when $s_{\max\text{-}verify}$ was high. These outcomes contradict Hypotheses 4a and 4b, showing that anchoring interacts with verification behavior rather than acting independently. This reflects empirical findings that people often anchor their beliefs to initial impressions, and that corrective interventions are more effective when trust in verification sources is already high.

In summary, susceptibility and verification thresholds are key factors driving misinformation dynamics, while anchoring bias introduces nonlinear effects that depend on the verification rate. These results underscore that interventions must be tailored to the cognitive and behavioral traits of their audience. While the model abstracts from real-world complexity, it captures essential dynamics that mirror how misinformation spreads and persists in digital environments.

## Discussion

This research provided interesting insights into the influence of individual behaviour on the diffusion of misinformation. The most important next step would be to verify this behaviour with real-world data. There have been efforts to observe information spreading in online communities (Knuutila, Herasimenko, Au, Bright, and Howard (2021), Memon and Carley (2020)). Correlating those results with our model would provide a reference for the model's accuracy.

Similarly, the uniform distribution of our susceptibility parameter could also be improved by having a stronger basis in observed real-world values. One could hypothesize that a normal-distributed value with a mean of 0.5 is more likely to be realistic, where 'ultra-susceptible' or 'ultra-sceptic' individuals are relatively rare.

For further research, we suggest an additional metric that represents the highest peak of number of believers he hoax has reached. Our current method only considers the time at which the peak of believers is reached, and the amount of believers at the end of the simulation. Thus these metrics might only paint a limited picture with more insights to be gained.

Additionally, performing separate measurements for various of the model's behavioral rules would also be helpful to better attribute the effect we are observing to specific changes. The enhanced model has the susceptibility parameter inform various simulation rules (As detailed in subsection Enhanced model). Investigating these rules individually could provide more insights into how they impact the simulation.

Finally, the results from this study could be further linked to efforts to prevent spreading of misinformation. The results of the various effects in the simulation could inform which individual behaviors would influence the spread of misinformation the most. For example: In our model, countering the effect of anchoring bias would be an effective strategy. Once validated against real-world data, these insights could strengthen interventions in the war on misinformation and achieve a positive impact.

## References

Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H., & Quattrociocchi, W. (2016, 06). Modeling confirmation bias and polarization. In (Vol. 7). doi: 10.1038/srep40391

Eurostat. (2021). *Evaluating data, information and digital content (2021 onwards)* (Tech. Rep.). European Commission. `https://doi.org/10.2908/ISOC_SK_EDIC_I21`. Retrieved from `https://ec.europa.eu/eurostat/databrowser/view/isoc_sk_edic_i21/default/table?lang=en`

Friggeri, A., Adamic, L., Eckles, D., & Cheng, J. (2014, May). Rumor cascades. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 101-110. Retrieved from `https://ojs.aaai.org/index.php/ICWSM/article/view/14559` doi: 10.1609/icwsm.v8i1.14559

Jost, P. J., Pünder, J., & Schulze-Lohoff, I. (2020). Fake news - does perception matter more than the truth? *Journal of Behavioral and Experimental Economics*, *85*, 101513. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2214804319303398` doi: https://doi.org/10.1016/j.socec.2020.101513

Knuutila, A., Herasimenko, A., Au, H., Bright, J., & Howard, P. N. (2021). A dataset of covid-related misinformation videos and their spread on social media. *Journal of Open Humanities Data*, *7*.

Li, J., & Chang, X. (2022, Jun). Combating misinformation by sharing the truth: a study on the spread of fact-checks on social media. *Inf Syst Front*, 1–15. (Epub ahead of print) doi: 10.1007/s10796-022-10296-z

Li, Z., Cao, J., Adams-Cohen, N., & Alvarez, R. (2023, November). The effect of misinformation intervention: Evidence from trump's tweets and the 2020 election. In *Disinformation in open online media* (pp. 88–102). Springer Cham. doi: 10.1007/978-3-031-47896-3_7

Memon, S. A., & Carley, K. M. (2020). Characterizing covid-19 misinformation communities using a novel twitter dataset. *arXiv preprint arXiv:2008.00791*.

Micallef, N., He, B., Kumar, S., Ahamad, M., & Memon, N. (2020). *The role of the crowd in countering misinformation: A case study of the covid-19 infodemic.* Retrieved from `https://arxiv.org/abs/2011.05773`

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330. Retrieved from `https://link.springer.com/article/10.1007/s11109-010-9112-2`

Pennycook, G., & Rand, D. G. (2018). Fighting misinformation on social media using crowdsourced judgments of news sources. *Proceedings of the National Academy of Sciences*, *115*(47), 12143–12148. Retrieved from `https://www.pnas.org/doi/full/10.1073/pnas.1806781116`

Pew Research Center. (2021). *Social media use in 2021.* Retrieved from `https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/`

Sherif, M., Taub, D., & Hovland, C. I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of experimental psychology*, *55*(2), 150.

Shiri, S., Hatef, B., & Sahraei, H. (2016, 03). Why human react differently to the same sensory experiences: an emotion-cognition interaction. *The Neuroscience Journal of Shefaye Khatam*, *4*. doi: 10.18869/acadpub.shefa.4.1.67

Sulis, E., & Tambuscio, M. (2020). Simulation of misinformation spreading processes in social networks: an application with netlogo. In *2020 ieee 7th international conference on data science and advanced analytics (dsaa)* (p. 614-618). doi: 10.1109/DSAA49011.2020.00086

Tambuscio, M., Ruffo, G., Flammini, A., & Menczer, F. (2015). Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks. In *Proceedings of the 24th international conference on world wide web* (p. 977–982). New York, NY, USA: Association for

Computing Machinery. Retrieved from https://doi.org/10.1145/2740908.2742572 doi: 10.1145/2740908.2742572

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. Retrieved from https://www.science.org/doi/10.1126/science.aap9559

Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., ... Quattrociocchi, W. (2019). Misinformation spreading on facebook. *EPJ Data Science*, *8*(1), 1–16. doi: https://doi.org/10.48550/arXiv.1706.09494