

DYNAMIC ANALYSIS OF FALSE INFORMATION SPREAD OVER SOCIAL
MEDIA: 5G-COVID 19 CONSPIRACY THEORY

by

Orkun İrsøy

B.S., Industrial Engineering, Boğaziçi University, 2019

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of

Master of Science

Graduate Program in Industrial Engineering

Boğaziçi University

2022

ABSTRACT

DYNAMIC ANALYSIS OF FALSE INFORMATION SPREAD OVER SOCIAL MEDIA: 5G-COVID 19 CONSPIRACY THEORY

The spread of false information via online social networks is a critical societal issue with various potential harms. Although there are huge efforts both in research and application to mitigate this problem, it persists with an increasing magnitude of results ranging from political manipulation to violent attacks. In our research, we built a causal simulation model to combine the existing accumulated knowledge in the literature and provide a formal model to evaluate the governing dynamics for the specific case of the viral spread of the 5G-COVID-19 conspiracy theory. The model makes use of both qualitative and quantitative data and successfully generates the observed dynamics for the 5g narrative. Results from the base run suggest that the dominance of believers in the active discussion on social media is overrepresented relative to the total population. Moreover, common mitigation strategies proposed in the literature such as limiting the interaction with believers of the misinformation often seem to produce worse outcomes for specific cases which indicates policy resistance. In addition, scenario analysis suggests that the involvement of neutral people in sharing misinformation or super-spreader actors might be enough to induce the system to pass the tipping point and generate an infodemic. The current analysis presents several trade-offs while discussing the underlying reasons through posterior analysis. In further research, we plan to expand our analysis by the inclusion of other user profiles, experiment with other mitigation strategies, and discuss the potential similarities and differences of our case with other types of false information dynamics.

1. INTRODUCTION

In today's world, communication methods have shifted significantly toward digital communication. The vast majority of people use social media, including people of all ages and socioeconomic backgrounds. As a result, an increasing number of people are using social media to gather and disseminate information on a variety of topics, including critical information. According to Reuters, nearly two-thirds of adult Americans use social media as a source of news (Moon, 2017).

This new mode of communication offers numerous benefits, including the promotion of engagement and the reduction of barriers among people all over the world by providing an alternative to face-to-face socialization. Information spreads faster and to a larger audience on these social networks. However, because the content is created by users without any review process, unlike traditional media, the content's validity cannot be verified. As a result, whether willingly or unwillingly, people may spread false information. Perhaps the most recent example demonstrating the potential harms of this phenomenon is the “infodemic” during the COVID-19 crisis, with results such as various ineffective and possibly harmful remedies, to outright rejection of the existence of the virus (Pennycook et al, 2020).

A recent example of such viral false information spread is 5G being one of the causes of COVID-19 or increasing its spread. The debate over the topic quickly erupted in the United Kingdom, particularly on social media platforms. Although fact-checking organizations or experts falsified the concerns related to this link, corrections were insufficient to alleviate the concerns, resulting in 5G tower arsons in Birmingham and Merseyside, United Kingdom (Ahmed et al., 2020).

Given the seriousness of the repercussions of misinformation dissemination, a massive body of research is carried out to tackle various facets of the problem. Researchers from various fields attempt to comprehend the psychological and cognitive drivers underlying the phenomenon, analyze the data at hand to deduce why and how false

information spreads, develop novel methods to detect such information, and develop mitigation strategies to combat misinformation.

Unfortunately, due to the complexity of the problem, mitigation measures used today are far from creating a structural solution but instead serve as symptom relief. The use of warning labels, which is one of the dominant tactics used by social media platforms, produces an "Implied Truth Effect" on unlabeled information (Penycook 2020) or may increase online traffic for the labeled content (Ingram, 2017). Fact-checking services attempt to verify the accuracy of the contents, although the rate of information production has increased far faster than the capacity of confirmation services has expanded (Penycook & Rand, 2019). Extensive data science research on false news detecting methods lays the door for the development of smart bots (Ammara, Bukhari, and Qadir, 2020).

Despite great efforts in both research and application, the failure of present mitigation techniques derives from the requirement for a dynamic systems approach. As a result, a systems perspective of the situation that combines current literature findings might identify potential leverage points and policy resistances to obtaining structural solutions to the problem at hand. In this regard, we argue that constructing a causal simulation model will constitute a theoretical basis to discuss the structural properties, derive practical insights, and experiment with different scenarios. As a specific case, we consider the aforementioned rumor spread regarding COVID-19 and 5G technology.

In the following sections, we provide a brief overview of both the general problem of misinformation spread and the COVID19-5G rumor, in particular, develop a System Dynamics model to gain insights into the problem. The credibility and the validity of the model and the resulting base behavior are discussed. Finally, we analyze some policy interventions and scenarios to infer some characteristics of the problem, and discuss the effectiveness of policies.

2. LITERATURE REVIEW & RESEARCH OBJECTIVES

Although false information dissemination is not a contemporary phenomenon, as evidenced by the 'Great Moon Hoax' in 1835 (Pennycook, Rand; 2021), the availability of highly linked worldwide platforms in the present world, allows anybody to transmit information to millions of individuals in a matter of minutes (Kumar and Shah, 2018) further increasing the reach and severity of the problem. False information causes issues ranging from political manipulation of large groups of people (Varol et al., 2017) and stifling rescue efforts during a crisis to even a terror strike (Kumar and Shah, 2018). One such instance is the Pizza Gate conspiracy, which resulted in a person firing a gun at a neighborhood shop in response to reports of child trafficking (Morstatter, Carley, and Liu, 2019). Another example is Facebook's claim of voter tampering in the 2016 Presidential Election (Lazer et al., 2018). Given the gravity of the effects, the World Economic Forum has identified false information spread on digital platforms as one of the main challenges to society (Lee Howell et al., 2013).

Many definitions and classifications of false information exist in the literature from rumors to "Fake News". One prevalent classification dimension is the intention of the agent where "misinformation" refers to unintentionally spreading information whereas "disinformation" is intentional (Wu et al., 2019; Kumar and Shah, 2018; Caled and Silva, 2021). Another categorization is the knowledge-based differentiation i.e., whether the information is purely factual or opinion-based (Kumar and Shah, 2018).

The problem of false information spread on social media has various drivers, both at the individual and aggregate levels. At the individual level, various psychological and cognitive factors are thought to be effective, and many researchers are trying to find answers to questions: Do political motives drive susceptibility to misinformation? Does repeated exposure lead to higher susceptibility to false beliefs? Which cognitive processes are influential on vulnerability to misinformation, and how can we design better corrective messages? (Penycook and Rand, 2021; Caled and Silva, 2021; Lewandowsky et al., 2021; Chan et al., 2017) From a more holistic perspective, another line of research focuses on the properties of these social networks, such as whether preferential attachment in these

networks forms echo chambers (repetitive exposure of specific information due to homogeneous social clusters), whether any structural reasons make specific networks more susceptible to misinformation spread, and whether any distinguishing characteristics differentiate the propagation dynamics of misinformation compared to true information (Vosoughi, Roy, and Aral, 2018; Vicario et al., 2016; Zhao et al., 2020). A huge effort is put into misinformation detection with machine learning using either content or context-based cues to design early interventions (Wu et al., 2019). Finally, simulation studies act as a testing platform to test the effectiveness of various intervention strategies or develop novel hypotheses about the underlying mechanisms of the problem (Kauk, Kreysa, and Schweinberger, 2021; Lotito, Zanella, Casari, 2021; Ammara, Bukhari, and Qadir, 2020).

Perhaps the most recent and critical forms of misinformation are experienced during the COVID-19 outbreak. Because of the ambiguity surrounding the situation, misleading information swiftly disseminates across borders, including conspiracy theories, fictitious miracle cures, and material that trivializes the infection (Bridgman, 2021). One such case that emerged in early January 2020 was the conspiracy theory suggesting a link between the installation of new 5G towers with the spread of the virus (Ahmed et al., 2020; Bruns, Harrington, and Hurcombe, 2020). Unfortunately, the spread of rumors did not solely become an instance of misinformation but the escalated panic yielded multiple attacks on 5G towers in the UK (Brewis, 2020; BBC, 2020).

Many researchers investigate different aspects of the “5G-COVID-19” conspiracy theory and its spread as it epitomizes the potential harms of viral misinformation. Ahmed and colleagues (2020) used Social Network Analysis to analyze the Twitter chatter during peak times of the chatter. Their analysis reveals that the number of people genuinely believing the conspiracy is rather low compared to the volume of the tweets. They conclude that apart from the believers, anti-conspiracy tweets, click baits or satiric tweets also contribute as much as believers, which further increases the dissemination of false information. Agley and Xiao (2021) conducted an online survey on the believability of various conspiracy theories including the 5G narrative. In their work, they analyze the relationship between believability scores for different profiles and their relationship with trust in science. Bruns, Harrington, and Hurcombe (2020) use both quantitative and

qualitative methods to understand how such misinformation escalated quickly up to violent attacks. They analyze the Facebook conversations from the start of the first rumor until the arson attempts by defining different phases and providing in-depth analysis for each phase. They discuss how pre-existing conspiracy networks or super-spreaders such as celebrities affected the propagation and the potential pitfalls that resulted in such virality. From a more quantitative perspective, Kauk, Kreysa, and Schweinberger (2021) use epidemiology modeling to build SIR (Susceptible-Infected-Recovered) model to simulate different mitigation strategies such as fact-checking and tweet deletion and evaluate their effectiveness. The authors also point out a few shortcomings of the SIR model such as reappearing incidence bursts and extreme peaks observed and call for more complex models that account for such behavior.

Given the severity and complexity of the problem at hand, the huge magnitude of the literature on misinformation is reasonable. However, the current attempts to combat misinformation are far from being effective. Since the research in this domain usually focused on one specific dimension of the problem such as propagation, detection, psychological factors, or network properties; the holistic view of the problem is yet to be achieved. In this regard, we argue that developing a formal dynamic simulation model will help to i) identify the causal feedback structure to gain insights into governing dynamics, ii) evaluate the effectiveness of potential structural mitigation strategies, and iii) discuss the similarities and disparities of the general structure for different cases of misinformation. System Dynamics methodology is an ideal fit for such a task as it allows the integration of main dynamic factors and provides a causal interpretation of the emerging dynamics. To narrow it down, we consider the 5G-COVID-19 conspiracy theory as exemplary. We believe that from a methodological perspective the constructed model will serve as a roadmap of potential improvements of SIR models of information diffusion to account for more complex cases, and from the perspective of the problem domain, it will contribute to a deeper understanding of the problem and provide a testing environment for different policies.

3. METHODOLOGY

In this study, the System Dynamics methodology is utilized to model the problem at hand. It is a widely used modeling methodology that is particularly used to address complex problems with many interacting components. System Dynamics models provide a dynamic hypothesis about the causal structure that generates the problem and allow the analysis of various policies and scenarios using simulation (Sterman, 2000).

The problem addressed in our study is dynamic in nature, includes a human dimension, and contains complex interactions, feedback loops, and nonlinearities which makes System Dynamics particularly useful since it is advantageous in addressing such properties (Barlas, 2002).

The main building blocks of System Dynamics models are “Stocks” and “Flows”. Stocks represent the accumulated variables which can be either physical (Body Weight) or abstract (Stress Level) whereas flows represent the changes in stock variables. Inflows (Weight Gain) increase the stock level whereas outflows (Weight Loss) decrease it. An additional variable type is auxiliary variables or converters that are used to denote the intermediary parameters included in the model boundary. They can be either constant during the simulation or they can be dynamic either defined by an equation or a graphical function.

Stock-Flow Diagrams are an adequate form of representation of the system being analyzed. One such diagram for illustration purposes is presented in Figure 3.1. The model is a slightly changed version of the population model presented by Barlas (2002). Here *Population* is a stock variable with the inflow of *Births* and outflow of *Deaths*. The Auxiliary variables are *Birth Fraction* and *Average Lifetime*. The mathematical relationship between stocks and flows is based on differential equations (3.1).

$$\text{Population}(t) = \text{Population}(t - dt) + (\text{Births} - \text{Deaths}) * dt \quad (3.1)$$

$$\text{Births} = \text{Birth Fraction} * \text{Population}$$

(3.2)

$$Deaths = \frac{1}{Average\ Lifetime} * Population \quad (3.3)$$

In Figure 3.1, in addition to the causal links represented, the polarity information of such links are also presented. The polarity of a causal link indicates the possible effect of a change in the causing variable on the affected variable. More clearly, a positive sign indicates that an increase (decrease) in the causing variable would result in an increase (decrease) in the affected variable's value as compared to the value that it would take if there were no change in the beginning. Conversely, a negative sign indicates that an increase (decrease) in the causing variable would cause a decrease (increase) in the affected variable as compared to the value that it would take otherwise.

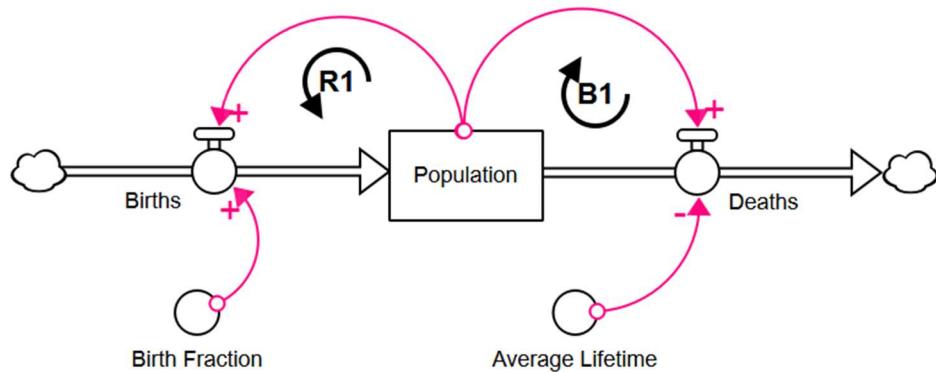


Figure 3.1: Stok-flow Diagram for the simple population model with causal loops

Along with the causal information, incorporating polarities of causal links also reveals the underlying causal loops which can be either balancing or reinforcing. The polarity of a loop is determined by multiplying the polarities of each individual causal link along the path. If the result of the calculation is 1 then the loop is a “reinforcing” loop or if it is -1 then the loop is “balancing”. Identifying such loops in the causal structure allows us to infer structural reasons for resulting behavior as we can define behavioral modes for reinforcing or balancing loops. Reinforcing loops usually generate either increasing growth or collapse whereas balancing loops often generate goal-seeking behavior.

In Figure 3.1, R1 represents the reinforcing loop created by the *Population* and *Births*. An increase in the *Population* would result in more *Births* which also causes an increase in the *Population*. Conversely, a decrease in *Population* would cause fewer *Births* which results in a decrease in the level of *Population* as compared to the value that it would take if there were no decrease in *Births*. On the other hand, B1 represents the balancing loop created by *Population* and *Deaths*. As *Population* increases, it causes more *Deaths* which balances the *Population* in turn.

In addition to modeling tools mentioned above, System Dynamics methodology will allow us to simulate the model at hand using differential equations. Such quantitative analysis is not a strict necessity as merely providing a dynamic hypothesis about the problem at hand, and identifying causal structures might also provide useful insights. However, as we applied in our study, using quantitative simulation for scenario and policy analysis is also useful in terms of generating quantitative comparisons and deductions. As a prerequisite of the credibility of such an analysis, both structure and behavior of the model should be tested, required parameters should be estimated using both quantitative and qualitative data in the literature prior to run simulation experiments. After the validity of the model is constructed, then analysis can be continued for the problem and further analysis can be applied for various scenarios and policy interventions.

Following the fundamental steps of System Dynamics methodology suggested by Barlas (2002); thus far we provided basic information about the issue and identified the problem. In the following sections, the formal model and a dynamic hypothesis is provided using differential equations, stock-flow & causal loop diagrams. Moreover, using the data at hand parameters are estimated and the validity of the model is discussed. Finally, the base behavior, various scenarios and policy interventions are analyzed, and conclusions are discussed in the last section.

4. OVERVIEW OF THE MODEL

A simplified version of the stock-flow diagram is presented in Figure 4.1. Fundamentally model is an enriched version of the traditional SEIR (Susceptible- Exposed- Infected- Recovered) model of information diffusion. Misinformation spread is often followed with corrective information either initiated by the disbelievers or by different authorities such as fact-checking organizations, scientific institutions, experts, etc. Thus, to differentiate between such a difference and reveal the competing dynamics between these groups, ‘Infected’ stocks are separated as “Believers” and “Disbelievers”. *Believer* stocks represent the people who think that the misinformation is true whereas *Disbeliever* stocks denote the people who are educated enough to know the rumor is false.

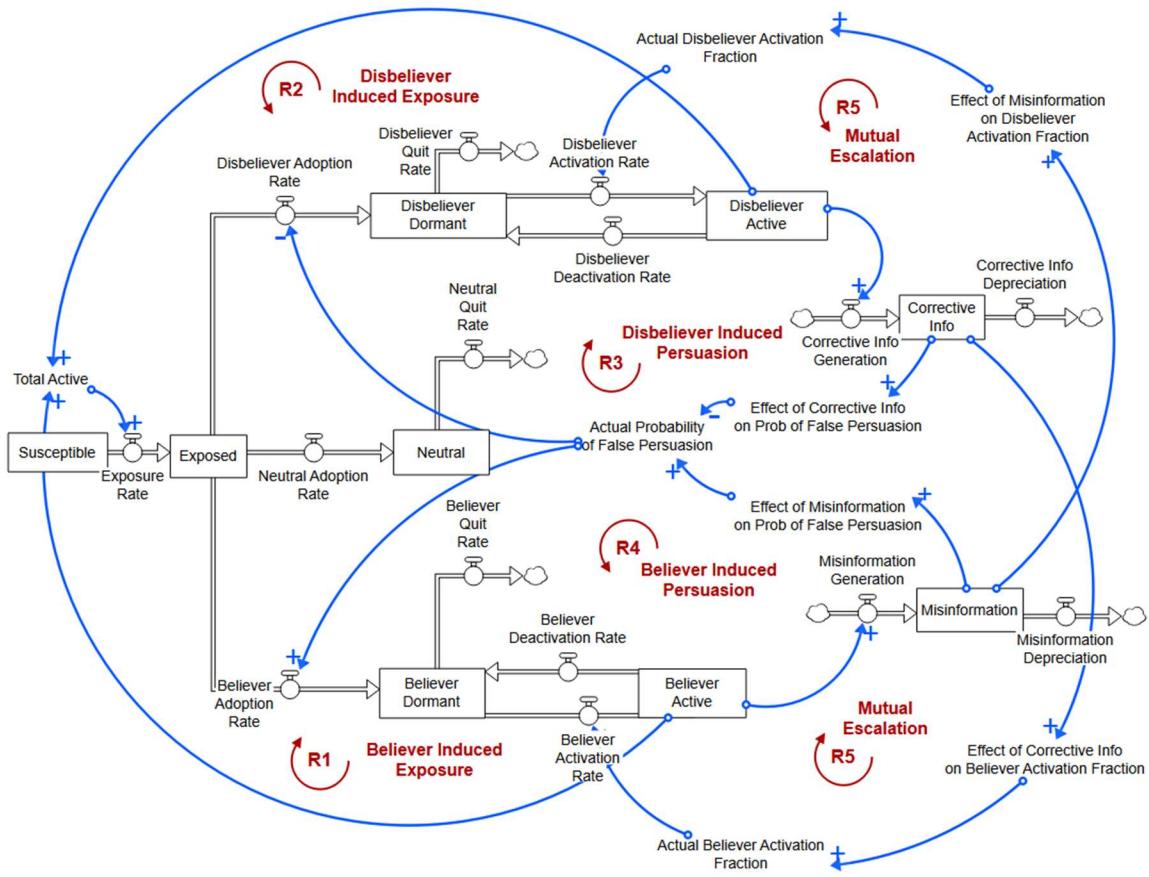


Figure 4.1: The simplified structure of the model and causal loops

Another distinction is made to differentiate between people who are actively spreading their views (either *Believer Active* or *Disbeliever Active*) on the issue or remain silent (*Believer Dormant*, *Disbeliever Dormant*, or *Neutral*). Therefore, *Believer Dormant* stock represents the people who believe the false information but remain silent on the issue whereas *Believer Active* is the people who believe and actively contribute to the spread of misinformation. Apart from *Believer* and *Disbeliever* stocks, people who are exposed can also remain neutral which is denoted as *Neutral*. The amount of information generated by *Believers* and *Disbelievers* is represented in *Misinformation* and *Corrective Info* stocks respectively.

The causal structure of the proposed model presents five main reinforcing loops:

Believer - Induced Exposure (R1) & Disbeliever - Induced Exposure (R2):

Exposure Rate is affected by amount of active people in the population. Thus, an increase in either *Disbeliever Active* or *Believer Active* stocks will result in an increased number of exposed people. Eventually, exposed people would proceed in this stock chain and increase the number of *Disbeliever* & *Believer Active*, closing the reinforcing loops.

Believer - Induced Persuasion (R3) & Disbeliever - Induced Persuasion (R4):

A constant fraction of *Exposed* becomes *Neutral (Recovered)*. The remaining fraction is split into *Believer Dormant* with *Actual Probability of False Persuasion* (p) and *Disbeliever Dormant* with the complementary probability ($1-p$). This probability is not constant as it is assumed that the available information would alter such fraction depending on the type of information. Thus, as the amount of *Misinformation* increases, the *Actual Probability of False Persuasion* also increases, resulting in more people adopting the false information. In turn, more believers would produce more misinformation closing the vicious cycle. A symmetric causal loop is present for the disbelievers as the increment in the *Corrective Info* would result in a smaller *Actual Probability of False Persuasion* thus increasing *Disbeliever Adoption Rate*.

Mutual Escalation (R5): A more complex loop emerges from the competing dynamics between the opposing groups. As *Misinformation* increases, more dormant disbelievers are inclined to share their opinion followed by an increase in the *Corrective*

info. Similarly, an increase in the Corrective info would result in more people becoming active for Believers. Such behavior is also reported in experiments conducted on digital social networks (Ma and Zhang, 2021).

5. MODEL DESCRIPTION

5.1. Stock-Flow Diagrams & Effect Formulations

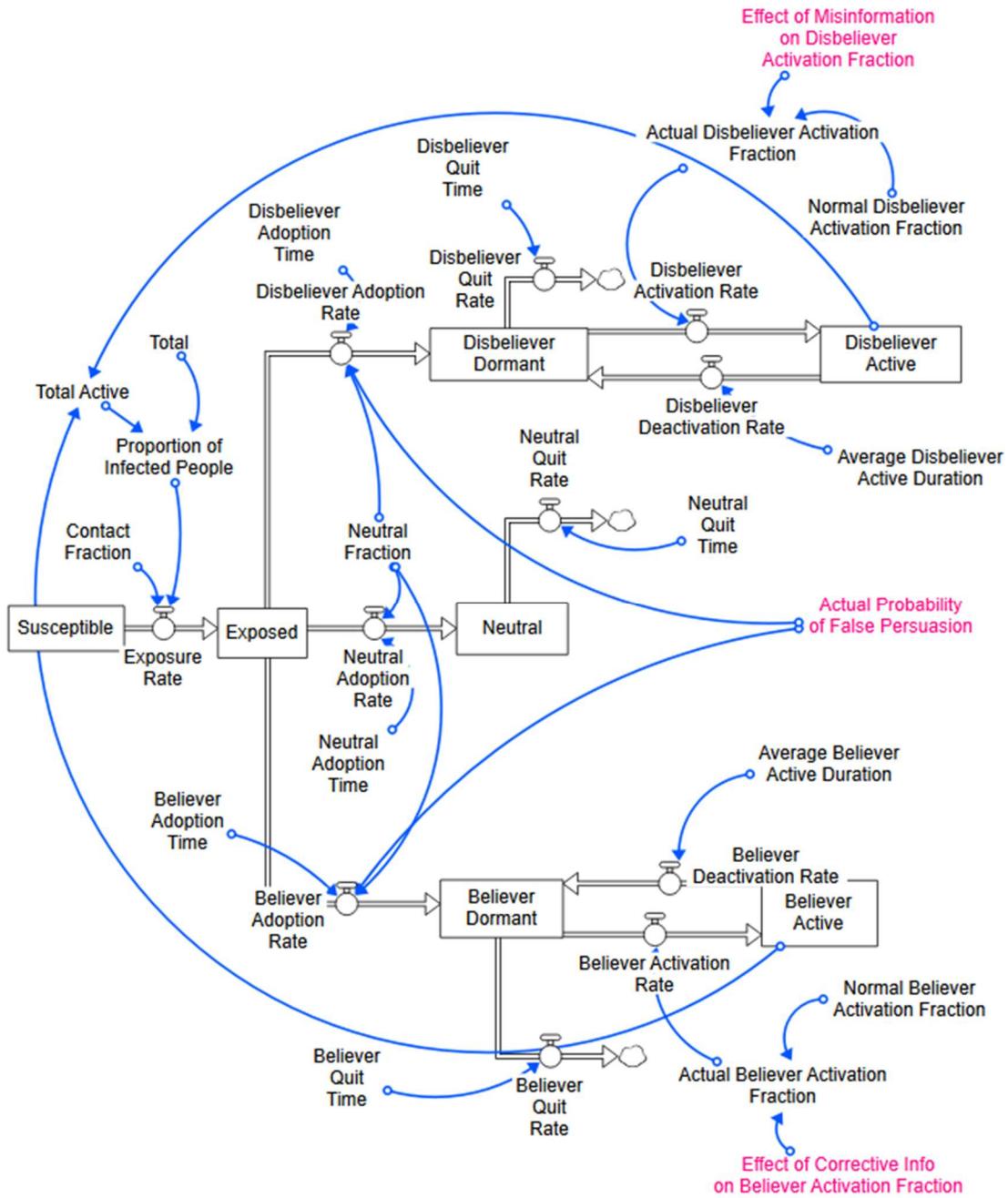


Figure 5.1. Population stocks (variables directly affected by *Misinformation* and *Corrective Info* are colored)

Figure 5.1 depicts the stock-flow structure for population stocks. The variables that are affected by *Misinformation* and *Corrective Info* are presented in a different color. *Susceptible* people contact other people with the *Contact Fraction*. The probability of such contact being with an active person is calculated by the ratio of *Total Active* to the *Total* number of people. Therefore, the resulting equation for *Exposure Rate* (5.1) is analogous to traditional SIR models with “Infected” people including both *Believer Active* and *Disbeliever Active*.

$$\text{Exposure Rate} = \text{Susceptible} * \text{Contact Fraction} * \text{Proportion of Infected People}$$

(5.1)

A constant fraction (*Neutral Fraction*) of *Exposed* remains neutral after the first exposure whereas the remaining is split between the opposing groups. The distribution is calculated using *Actual Probability of False Persuasion*. *Adoption Rate*, *Quit Rate*, and *Deactivation Rate* flows are modeled as typical delay formulations using average times as delays (5.2) whereas *Activation Rate* flows are formulated as a fraction of corresponding stocks using actual values of the fractions (5.3). Activation in each group is determined by the actual activation fractions which are multiplications of graphical effect functions with normal values.

$$\text{Outflow} = \text{Stock} / \text{AverageTime}$$

(5.2)

$$\text{Outflow} = \text{Stock} * \text{Fraction}$$

(5.3)

Figure 5.2 represents the information stocks and related effect formulations. *Misinformation Generation* and *Corrective Info Generation* are calculated by the multiplication of active stocks with the *Average Information Generation per person*. Generated information depreciates with the depreciation delays (5.2). All effect functions

are standardized using *Standard Misinformation per capita* & *Standard Corrective Info per capita*. Actual values for *Activation Fractions* are derived by multiplying effects with the normal values of the parameters employing standard multiplicative effect formulation (5.4).

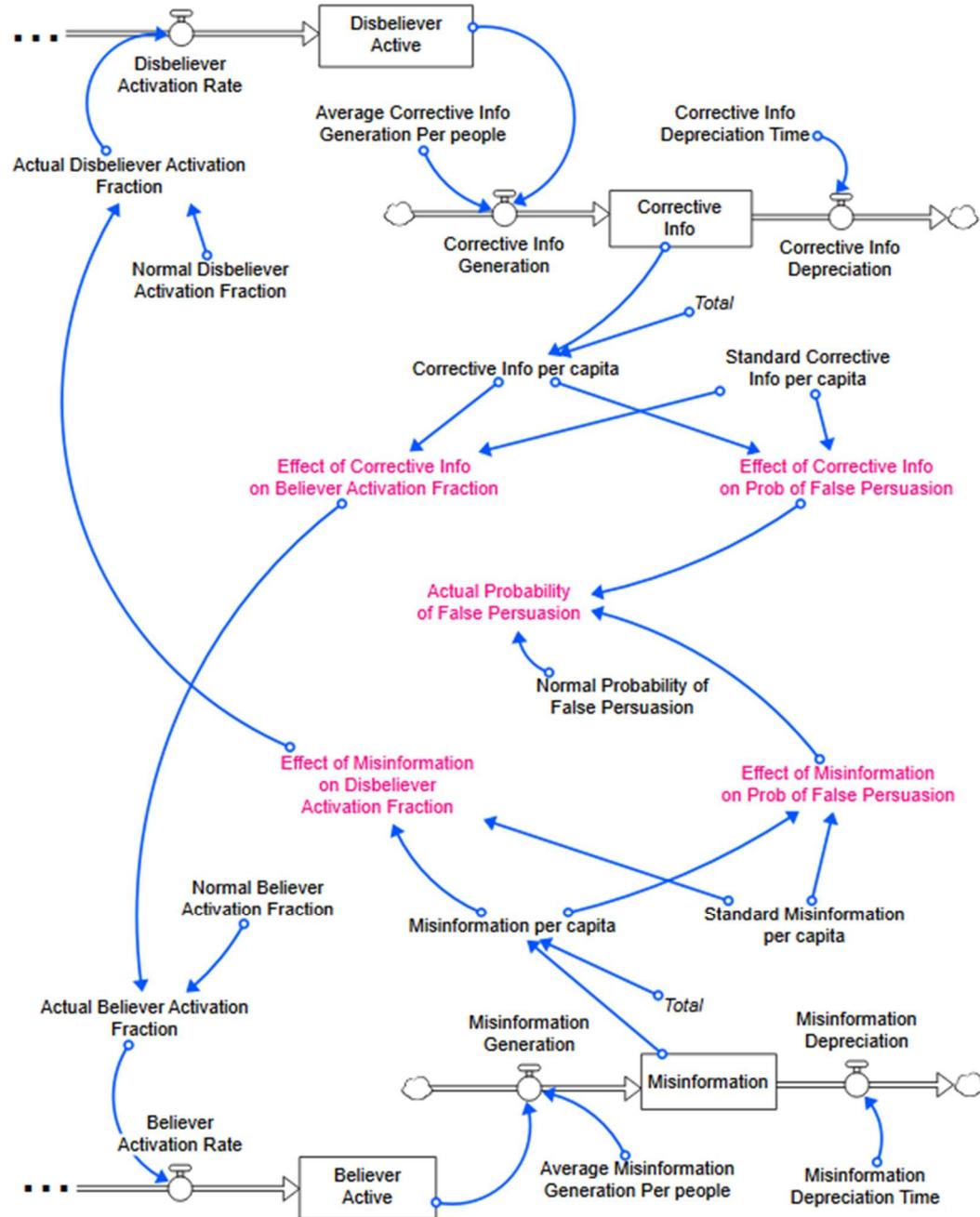


Figure 5.2. Information stocks and effect functions (variables directly affected by *Misinformation* and *Corrective Info* are colored)

Effect of Misinformation on Disbeliever Activation Fraction

$$= f\left(\frac{\text{Misinformation per capita}}{\text{Standard Misinformation per capita}}\right) \quad (5.4)$$

Actual Disbeliever Activation Fraction

$$\begin{aligned} &= \text{Normal Disbeliever Activation Fraction} \\ &\quad * \text{Effect of Misinformation on Disbeliever Activation Fraction} \end{aligned}$$

(5.5)

There are four graphical functions utilized in the model: two effects regarding the *Actual Probability of False Persuasion* and the other two regarding the activation fractions. Defined graphical effect functions of *Effect of Misinformation on Disbeliever Activation Fraction* and *Effect of Corrective Info on Believer Activation Fraction* are presented in Figure 5.3. For both, the input of the graphical function is the ratio of per capita values to their standardized values. A recent study reports that the decrease in perceived peer support increases opinion expression in the digital sphere, WhatsApp groups in particular (Ma and Zhang, 2021). Moreover, it is reported that sharing behavior of such misinformation is often affected by identity-based thinking (Oyserman & Dawson, 2020). Additionally for a disbeliever to become active and start generating corrective information, encountering the misinformation is a necessity. Thus, it is fair to assume that the *Effect of Misinformation on Disbeliever Activation Fraction* should be an increasing function of *Misinformation per capita*. Therefore, it is assumed that initially if misinformation is not present, disbelievers should stay in the dormant state whereas as *Misinformation per capita* reaches a theoretical standard point then the *Actual Disbeliever Activation Fraction* assumes its normal value (at point (1,1)). As *Misinformation per capita* passes beyond that standard point; disbelievers are getting more active as they encounter *Misinformation* more frequently. Such an effect should reach saturation level as the remaining people in the dormant

informed stock will be the least motivated ones to speak up. Thus, the increasing function is modeled as having S-shape.

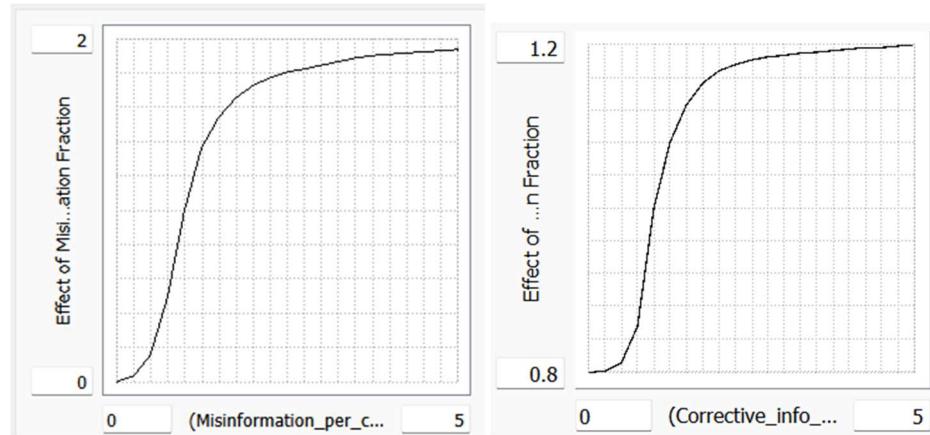


Figure 5.3: Graphical effect functions of *Effect of Misinformation on Disbeliever Activation Fraction* (left) and *Effect of Corrective Info on Believer Activation Fraction* (right)

The backfire effect of corrective action on misinformed people is somewhat contradictory. Some studies suggest that individuals might share information as an identity signaling mechanism especially people with conspiracy thinking. Thus, corrective information might be perceived as an attack on self-identity which might increase the activation. However, some other studies discuss that such backfire effects might be evaded if the corrective actions are designed in an effective way (see Ecker et al., 2022 for a detailed discussion). Nonetheless, for our case, we assume no transition between *Believer* & *Disbeliever Stocks*, thus only interested in whether such exposure to corrective information would cause any change in the sharing behavior of current believers. Given our specific case 5G-COVID 19, believers involve mostly conspiracy susceptible people. Consequently, we assumed that social group effects would be more prominent as the literature suggests for conspiracy thinkers (Prooijen & Douglas, 2018). Therefore, the *Effect of Corrective Information on Believer Activation Fraction* is modeled with the same logic as the misinformation effect, i.e. the effect is a logistic increasing function of *Corrective Information per capita*, only differing in minimum-maximum values. Considering the contradicting results in the literature, it is assumed that the effect should be less dominant compared to the *Effect of Misinformation on Disbeliever Activation Fraction*. Furthermore, since the motivation of believers' sharing behavior is not

necessarily driven by the existence of *Corrective Info* the effect function takes a non-zero value when the *Corrective Info* was zero.

The effects regarding the information effects on the *Probability of False Persuasion* are formulated using additive graphical effect functions. Thus, the *Actual Probability of False Persuasion* is calculated as the sum of *Normal Probability of False Persuasion*, *Effect of Misinformation on Prob of False Persuasion*, and *Effect of Corrective Info on Prob of False Persuasion*. The fact that repeated exposure contributes to the belief is a constructed phenomenon (Ecker, 2022; Dechêne et al., 2010; Hasher, Goldstein, & Toppino, 1977). Thus, the *Effect of Misinformation on Prob of False Persuasion* is an increasing function of *Misinformation per capita* since we would expect more people to be convinced by the misinformation on average as there is more misinformation available (Figure 5.4). Conversely, if competing *Corrective Info* is present, the believability of the false information should decrease, hence the corresponding effect is a decreasing function of *Corrective Info per capita*. The initial limits for these graphical functions are taken as 0.1 and -0.1 as a simplification.

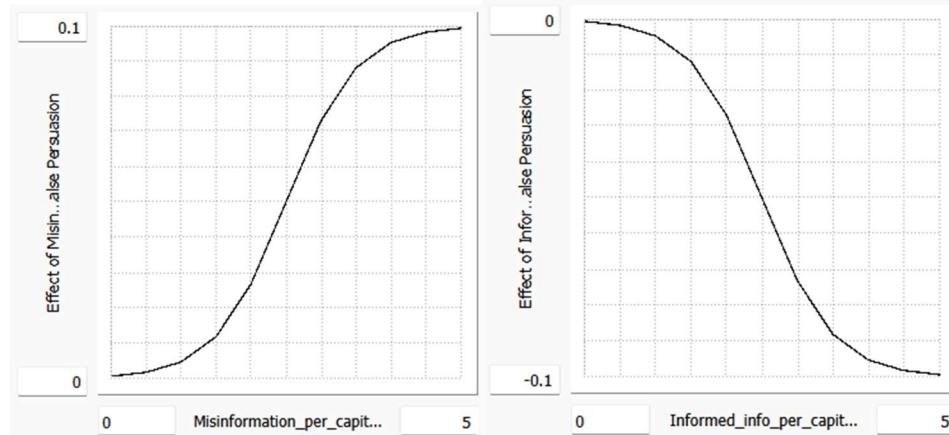


Figure 5.4. Graphical effect functions for *Effect of Misinformation on Prob of False Persuasion* (left) and *Effect of Corrective Info on Prob of False Persuasion* (right)

Finally, some additional clarifications on model assumptions & simplifications:

- Only active stocks (*Believer Active* and *Disbeliever Active*) contribute to the *Exposure Rate* for the base model. Thus, if there are no active people in both groups there wouldn't be any propagation even though there are dormant people.
- False persuasion probability (or believability of misinformation) is assumed to be independent of the exposure pathway (whether exposed by a disbeliever or believer). Thus, instead of using separate stocks for “Exposed by Believer” and “Exposed by Disbeliever”, these two stocks are aggregated in the *Exposed* stock.
- Since the *Exposure Rate* formulation is based on the contacts between *Susceptible* and *Active Stocks* (and not affected by *Misinformation* or *Corrective Information* directly) in case of no misinformation, there can still be propagation in this specific social media platform. Therefore the corresponding assumption is that even though we silence all misinformation on a specific platform, there will still be some people that encounter the information from other sources or social media platforms thus still causing a propagation but not contributing to chatter on this platform.
- *Normal Probability of False Persuasion* is assumed to have some constant value. The actual value dynamically changes depending on existing information ecosystem which implicitly assumes that there should be some fraction of the total population who can either believe or disbelieve based on the availability of the competing information. Thus, the range of *Actual Probability of False Persuasion* presents an estimate of the fraction of people that can change their minds based on whether there is competing information or not.
- *Believer Active Duration* and *Believer Activation Fraction* are assumed to be larger than the *Disbeliever Active Duration* and *Disbeliever Activation Fraction* respectively, as people having a conspiracy mindset have a larger tendency to insist on their viewpoint since they are personally involved in the issue.
- *Misinformation* and *Corrective Information* have an artificial unit of “information” instead of tweets or posts. The reason is that since they are used in effect formulations, they should represent the effect of that type of information which

should depreciate as time progress. Thus, rather than the physical unit of expression of that specific social media platform, these stocks are defined as soft variables and they correspond to sustained effects of their corresponding information type.

- The possibility of transition from believer to disbeliever (or vice versa) is not allowed, as well as the transition from *Neutral Stock* to any other stocks.
- All other variables except the ones with effect function formulations (*Activation Fractions* and *Probability of False Persuasion*) are assumed to be constant during the simulation horizon in the base model, although the sensitivity results for each one are presented as supplementary material.

6. PARAMETER ESTIMATION AND VALIDITY OF THE MODEL

6.1. Parameter Estimation

The initial parametrization that constitutes “Base Run” is provided in Table 6.1. To deduce the model parameters, various research from the literature is utilized. Based on the believability scores obtained in the study of Agley and Xiao (2021), the *Normal Probability of False Persuasion* is kept at around 0.2 during calibration. Initial stock values of *Misinformation* and *Corrective Info* are assumed zero as the spread is assumed to start at time t=0. Since we are mainly concerned with the dynamics of the system rather than acquiring a perfect fit to data, we select the total number of people in the system as 10000-a close number (approx. 9999) of the total estimated in Kauk, Kreysa, and Schweinberger, (2021)-, assume the initial level of 10 for *Believer Active* to start the propagation, and assume the initial value of 0 for the other stocks.

Parameter Name	Unit	Value	Stock Name	Unit	Initial Value
Normal Prob of False Persuasion	-	0.22 [1]	Believer Active	person	10
Neutral Fract	-	0.1 [3]	Believer Dormant	person	0
Contact Fraction	day ⁻¹	0.8 [3]	Exposed	person	0
Believer Quit Time	day	9.09 [2]	Disbeliever Active	person	0
Disbeliever Quit Time	day	9.09 [2]	Disbeliever Dormant	person	0
Neutral Quit Time	day	9.09 [2]	Corrective Info	information	0
Average Believer Active Duration	day	3 [3]	Misinformation	information	0
Average Disbeliever Active Duration	day	1 [3]	Neutral	person	0
Normal Believer Activation Fraction	day ⁻¹	0.68 [3]	Susceptible	person	10000
Normal Disbeliever Activation Fraction	day ⁻¹	0.2 [3]			
Average Corrective Info Generation per people	information/(day*person)	1 [3]			
Average Misinformation Generation per people	information/(day*person)	1 [3]			
Corrective info Depreciation Time	day	2 [3]			
Misinformation Depreciation Time	day	2 [3]			
Standard corrective info per capita	information/person	0.02 [3]			
Standard misinformation per capita	information/person	0.02 [3]			

Table 6.1. Parameter values and initial levels of stocks. [1]: Agley and Xiao, 2021; [2]: Kauk, Kreysa, and Schweinberger, 2021; [3]: Calibrated using data from: Ahmed et al., 2020; Kauk, Kreysa, and Schweinberger, 2021.

The analysis conducted by Ahmed and colleagues (2020) revealed that the prevalence of pro-conspiracy (34.8%) and anti-conspiracy (32.2%) tweets about the issue

is quite close for the 1 week during the peak time of the debate. Thus, the calibration is made so that the *Believer Active* should be slightly larger than *Disbeliever Active* during the peak of the chatter.

To calibrate the remaining parameters two datasets from recent studies (Kauk, Kreysa, and Schweinberger, 2021; Langguth et al., 2022) are utilized. In their work, Kauk, Kreysa, and Schweinberger (2021) use Twitter hashtag data to approximate the level of “Infected” people for the traditional SIR model. The hashtag data consist of the number of daily tweets containing specific hashtags, mostly pro-conspiracy hashtags such as “#stop5g”, “5gkills”, ”5gcoronavirus”, etc. Due to the volatility of daily hashtag data, the authors use the cumulative version of the data for the parameter calibration. The other study (Langguth et al., 2022) analyzes the long-term spread of the 5G-COVID 19 conspiracy with a specific emphasis on the spatial spread of misinformation. In their study, authors collect every tweet containing specific keywords such as “5G”, ”5g”, ”coronavirus”, ”COVID”, etc. for two years and then use manual labeling and machine learning to obtain an estimate of the exact number of tweets that contains misinformation. The authors provide the monthly aggregated version of the posted tweet that contains misinformation and is posted in the UK. Our simulation, however, spans the six months interval with time unit as days. Thus, the data in this study lack the necessary resolution to calibrate the model parameters. Therefore, we use the daily hashtag data from Kauk, Kreysa, and Schweinberger (2021) for numerical calibration and normalized monthly posted tweets data from Langguth et al. (2022) to check the behavioral characteristics.

Parallel to the model fitting process in Kauk, Kreysa, and Schweinberger (2021), the cumulative hashtag data is used for numerical calibration. The model at hand does not count the number of tweets explicitly. Therefore, to obtain a proxy for the cumulative incidence of tweets, it is assumed that an average active believer posts one tweet per week (*Average Time to Tweet* = 7 days) under these hashtags. Thus, the *Cumulative Believer Tweets* are defined as a stock with a constant inflow equal to *Believer Active / Average Time to Tweet*. Although a tweet per week might seem like a small frequency, since the data at hand is the number of tweets under specific hashtags rather than all tweets about the issue,

it is adequate to assume a longer period for *Average Time to Tweet* to avoid underestimating the number of believers.

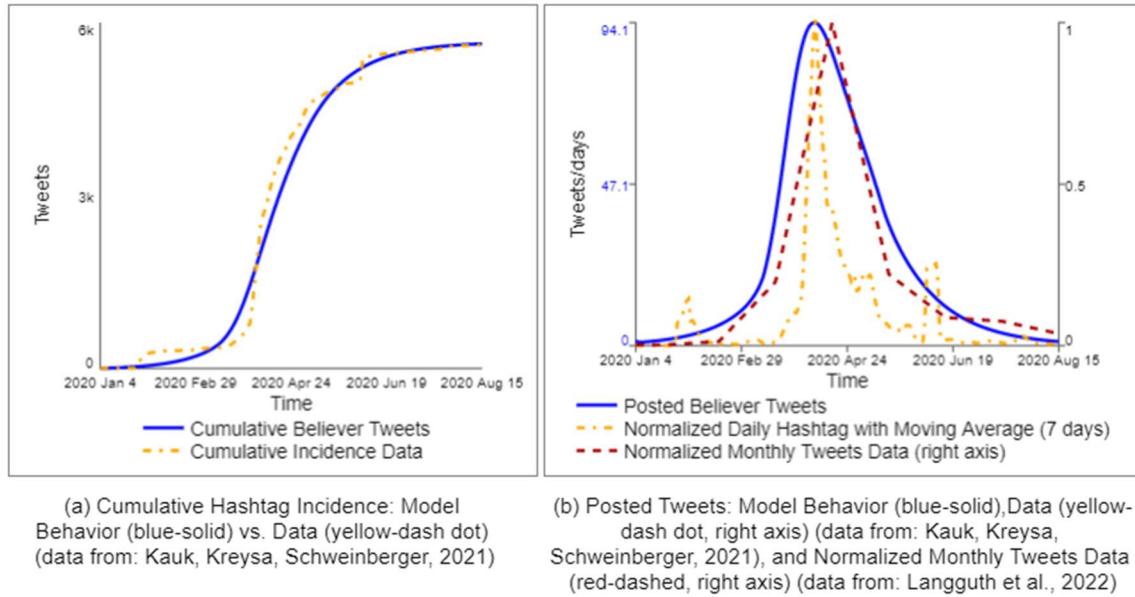


Figure 6.1. Model behavior after calibration vs. various data sources

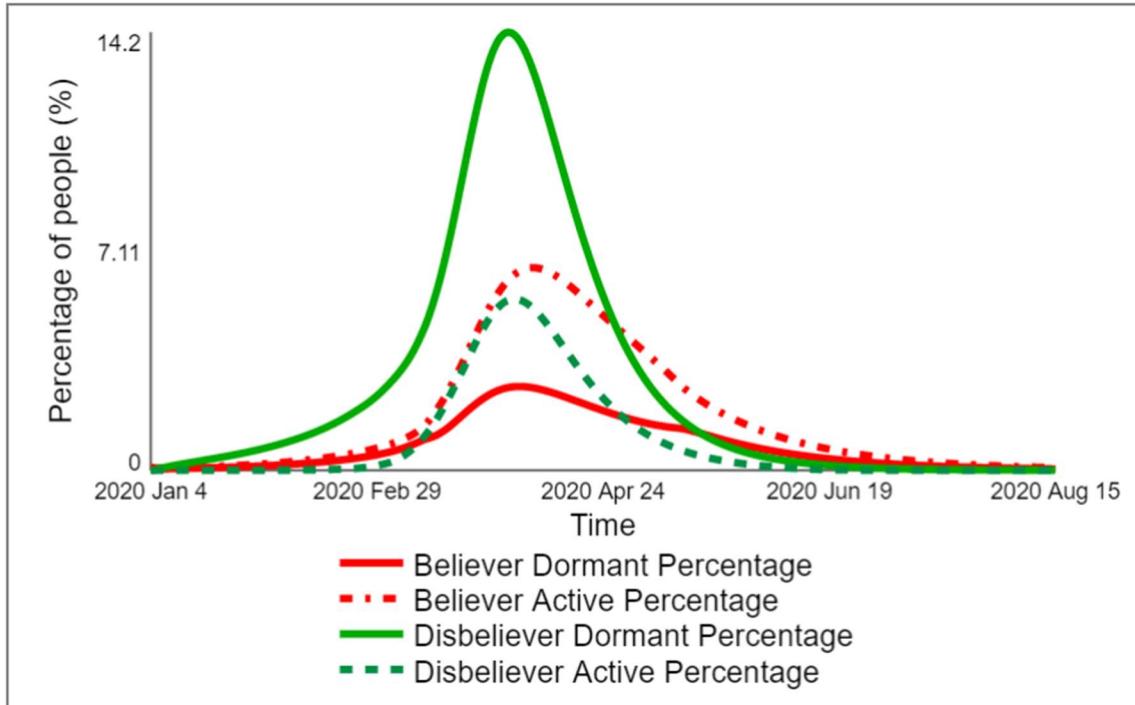


Figure 6.2. Dormant & Active Stocks after calibration

Figure 6.1. depicts the comparative graphs of data provided in the literature (dash-dot yellow lines) and the model behavior (solid blue lines). As provided in Figure 6.1.a., the resulting *Cumulative Believer Tweets* provide a good fit with the cumulative hashtag incidence data from Kauk, Kreysa, and Schweinberger (2021). *Posted Believer Tweets* peak around the 8th of April, 2020 which coincides with the peak time observed in daily hashtag data (Figure 6.1.b). Moreover, the *Believer Active* to *Disbeliever Active* ratio is close to 1 during the peak (Figure 6.2.) which is in line with the findings provided by Ahmed and colleagues (2020). Although the current set of parameters does not explain the reoccurring peaks in hashtag data (Figure 6.1.b), the base model is built with this parameter setting and whether such differences in the behavior can be obtained by further expanding the dynamic hypothesis or changing the parameter settings will be evaluated later.

6.2. Structural Credibility

Structural validity is an essential part of the modeling process, especially for the causally descriptive simulation models. To reveal the underlying structural reasons causing the problem at hand or to provide policy analysis with real-life implications, the foremost condition is that the model equations and other causal relations embedded in the model should be consistent with the available knowledge about the real system (Barlas, 1996). Regarding this, the model structure is built upon the existing qualitative and quantitative literature, utilized mathematical equations are criticized with direct structure tests, and parameter consistency is ensured along the model building process. All model parameters have meaningful real-life counterparts, and assumptions or simplifications are clarified in the “Model Description” section.

6.3. Structure - Oriented Behavior Tests

To further solidify the model validity, we applied several extreme condition tests as proposed by Barlas (1996). These tests aim to ensure that the simulation model would produce valid behaviors under extreme conditions as the real system would produce. Some examples of such conditions are having no initial active people, a very small probability of false persuasion, a very high probability of false persuasion, a sudden decrease in the *Misinformation*, or a sudden decrease in the *Believer Active*. The resulting behaviors for

the last two cases are presented in Figure 6.3 and Figure 6.4 whereas the results of the others are provided in Appendix B.

If the believability of misinformation is too low or if there are no people knowing that misinformation, one should expect no spread of such information. Conversely, if the believability of misinformation is too high, we would expect that misinformation spreads to most of the population without many disbelievers. To test the model behavior under these extreme conditions, three simulation experiments are conducted with three different parameter settings: *Normal Probability of False Persuasion* = 0.001 (low believability), *Normal Probability of False Persuasion* = 0.999 (high believability), *Initial Active* = 0 (no people knowing misinformation). The resulting behavior of the model (see Appendix B.) is consistent with the real-life expectation for the three cases.

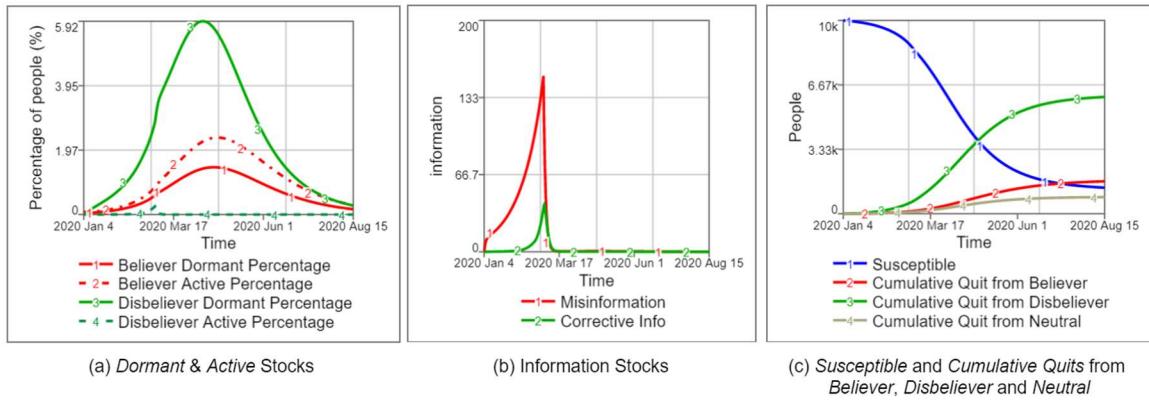


Figure 6.3. Model behavior for the extreme condition of eliminating *Misinformation*

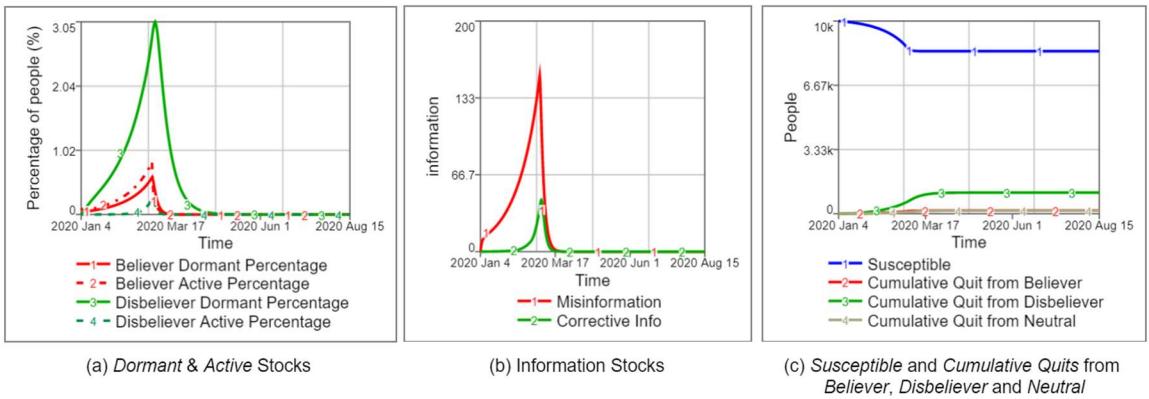


Figure 6.4. Model behavior for the extreme condition of eliminating *Believer Active*

The results of the other two tests are presented in Figures 6.3 and Figure 6.4. To test the behavior of the model, external outflows are activated on day 60 and implemented to *Misinformation* (Figure 6.3) and *Believer Active* (Figure 6.4) stocks. Elimination of *Misinformation* after day 60 (Figure 6.3.b) results in no dispute on the subject for this platform whereas the propagation still exists (Figure 6.3.a) with a lesser impact as the active people continue to spread the misinformation outside this social media platform. Elimination of *Believer Active* on day 60 (Figure 6.4) on the other hand, stops the propagation as observed in the stabilized levels in *Susceptible* (Figure 6.4.c), thus also ending the dispute in this specific social media platform (Figure 6.4.b).

7. BASE RUN AND SENSITIVITY ANALYSIS

7.1. Base Run

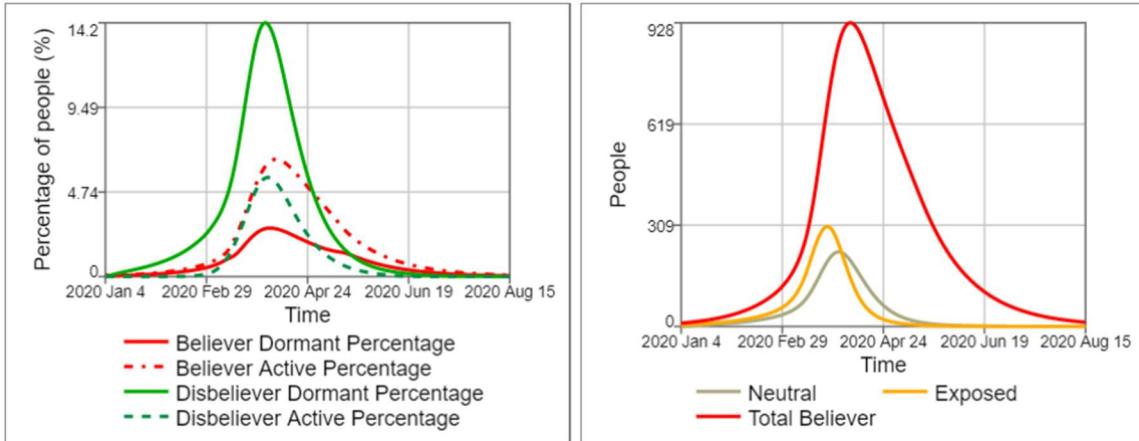
To compare the effectiveness of different runs, three outcomes of interest are defined: *Exposure Percentage*, *Total Believer Peak Percentage*, and *Believer Incidence Percentage*. *Exposure Percentage* denotes the percentage of the total population that have been exposed to this information, *Total Believer Peak Percentage* is the maximum number of total believers (both *Dormant* and *Active*) as a percentage of the whole population, and *Believer Incidence Percentage* is the percentage of the total population that has been a believer at least once in simulation horizon. For all three of them, we consider the final values at the end of the simulation run. The motivation behind defining different measures is that based on the nature of the misinformation, policymakers might target different outcomes. For instance, policymakers might try to minimize *Total Believer Peak Percentage* while countering misinformation that is expected to create violent responses in the audiences as it was in our specific case and numerous other ones (Atakan & Sen, 2022; Morstatter, Carley, and Liu, 2019). On the other hand, considering the stickiness of misinformation and its sustained effects even after the debunking (Ecker, 2022; Lewandowsky et al., 2012), policymakers might prefer minimizing *Believer Incidence Percentage* if the misinformation affects the long-term public health behavior, for instance.

The base run represents 225 days (7.5 months) from the 4th of Jan to the 15th of Aug where the debate is prominent on Twitter as hashtags. The dynamics of the main stocks and variables are presented in Figure 7.2 and resulting outcome of interests are presented in Figure 7.1. The behavior of the four people stocks (Figure 7.2.a) is similar, as the peak times and the shapes are nearly the same for all 4 stocks. *Misinformation* seems to exceed the *Corrective Info* for the whole period and nearly all of the *Susceptible* is depleted with Exposure Percentage equal to 99.22 %. Considering our assumption that *Susceptible* represents the people on the social media platform that has the potential to participate in the discussion, we can say that the maximum potential is reached for this case. It seems intuitively consistent, as the 5G narrative is one of the most viral instances of misinformation involving distribution channels such as national TV, celebrity super-

spreaders, and conspiracy theorists with preexisting social connections (Bruns, Harrington, and Hurcombe, 2020) thus resulting in a wider reach to various audiences.

Exposure Percentage	99.22%	Total Believer Peak Percentage	9.271%	Believer Incidence Percentage	20.13%
---------------------	--------	--------------------------------	--------	-------------------------------	--------

Figure 7.1. Resulting outcomes of interest for the Base Run



(a) Dormant & Active Stocks

(b) Neutral Dormant, Exposed, and Total Believers

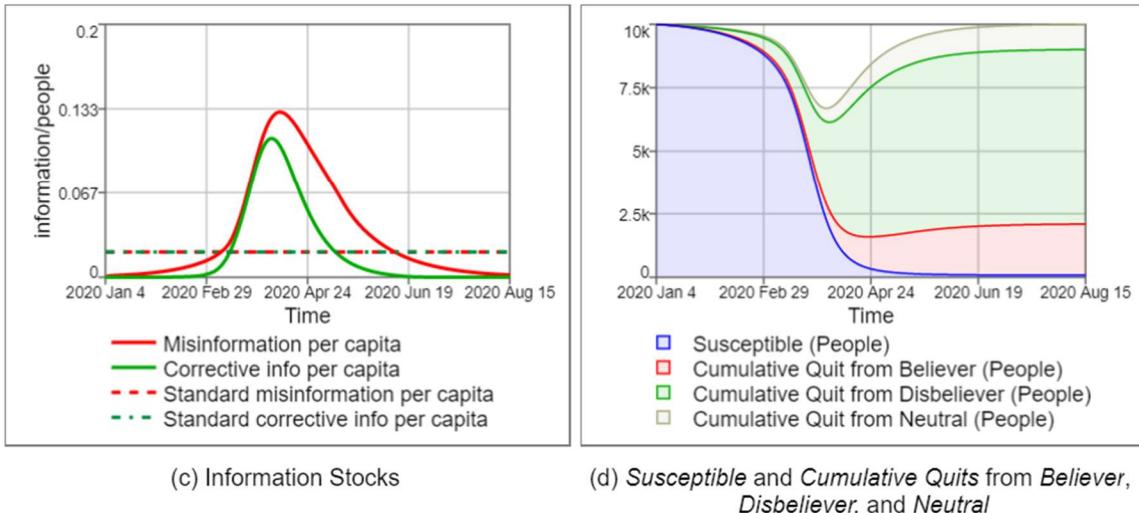


Figure 7.2. Resulting stock values for the Base Run

The initial observation in Figure 7.2.a is that although the magnitude difference between the *Believer Dormant* and *Disbeliever Dormant* is huge, the *Active Stock* levels are quite close for the two groups. Therefore, one simple insight is even when the pro-conspiracy people in the population are in minority, their presence in the digital sphere (i.e.

Active Stocks) can dominate the educated people, as believers are more inclined to engage in social media. It should be noted that such an insight is provided by the enriched model whereas the traditional SIR models lack the necessary resolution for such an analysis.

7.2. Sensitivity of Model Behavior to Changes in the Model Parameters

Sensitivity analysis is conducted to analyze the model behavior for changes in model parameters. To avoid combinatorial complexity, the analysis is conducted one parameter at a time. Table 7.1 presents the selected parameters for sensitivity analysis and the minimum and maximum values included in the analysis. Full results of the sensitivity analysis are presented in Appendix C whereas important runs are discussed in this section. The sensitivity runs are taken for 300 days (instead of 225 days in the Base Run) to be able to observe the long-term dynamics that might be caused by the changes in some of the parameters such as *Contact Fraction*.

Parameter Name	Unit	Value (Base Run)	Min -Max	Increment
Normal Prob of False Persuasion	-	0.22	0 - 0.5	0.1
Contact Fraction	day -1	0.8	0.1 - 0.9	0.1
Average Believer Active Duration	day	3	1 - 4	0.5
Average Disbeliever Active Duration	day	1	0.5 - 3	0.5
Normal Believer Activation Fraction	day -1	0.68	0 - 1	0.2
Normal Disbeliever Activation Fraction	day -1	0.2	0 - 0.5	0.1
Corrective info Depreciation Delay	day	2	1 - 3	0.5
Misinformation Depreciation Delay	day	2	1 - 3	0.5
Standard corrective info per capita	information/person	0.02	0.01 - 0.05	0.01
Standard misinformation per capita	information/person	0.02	0.01 - 0.05	0.01

Table 7.1. Selected Parameters for sensitivity analysis, units, base run values, analysis intervals, and increments

Overall, the model shows consistent behavior in terms of model assumptions and real-life expectations. The most influential variable in terms of changes in the outcome of interests is the *Normal Probability of False Persuasion* (Figure 7.3). Since the parameter itself represents the believability of the misinformation, the model behavior is consistent with the expectation that the more believable the false information is the more people would fall for it. Another factor to notice is that, considering the changes in *Exposure*

Percentage analogous to epidemiological models, there is a threshold value for *Normal Probability of False Persuasion* around 0.1 and 0.2 below which the epidemic does not start. If the believability of the information is above that threshold, *Believer Incidence Percentage* and *Total Believer Peak Percentage* seem to linearly increase with the increasing probability of believing.

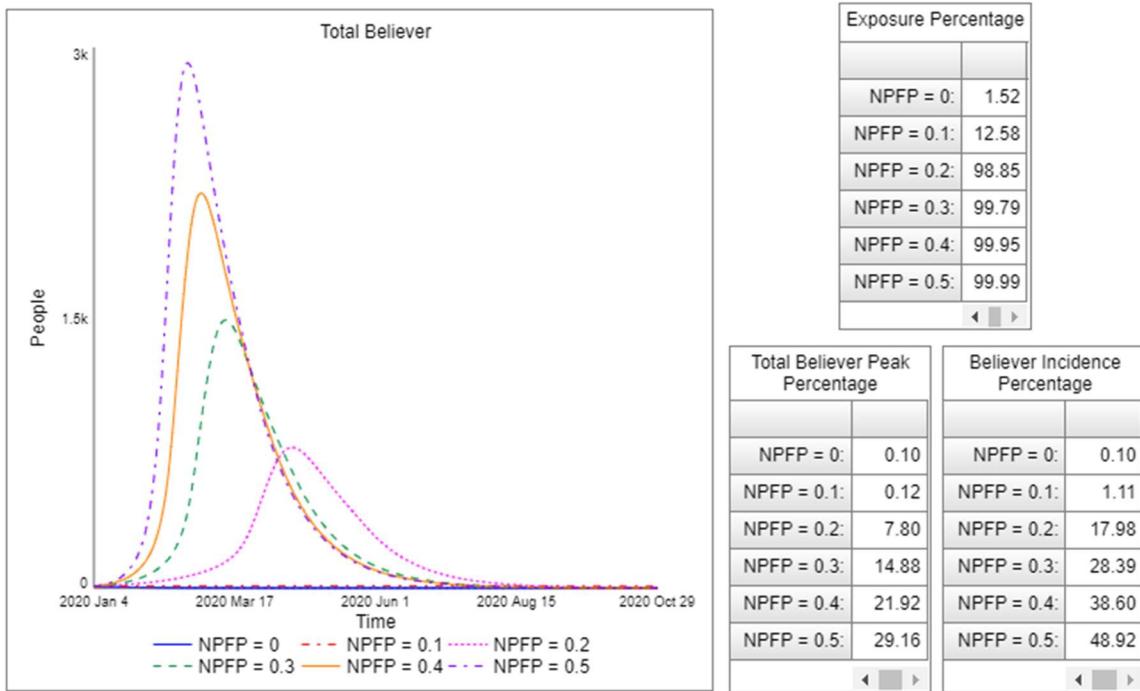


Figure 7.3. Sensitivity of *Total Believer* to *Normal Probability of False Persuasion (NPFP)*

The *Normal Probability of False Persuasion* is not the only parameter with a threshold property. Figure 7.4 depicts the *Susceptible* levels for different values of *Contact Fraction*, *Normal Believer Activation Fraction*, and *Average Believer Active Duration*. We observe none to minuscule changes in *Susceptible* levels for *Contact Fraction* below 0.3 (Figure 7.4.a), *Normal Believer Activation Fraction* (Figure 7.4.c) below 0.2, and *Average Believer Active Duration* below 1 (Figure 7.4.d). These four parameters are the only parameters that result in such behavior in the scope of this sensitivity analysis (see Table 7.1). *Normal Probability of False Persuasion* and *Contact Fraction* are related to the different characteristics of the misinformation and the structural properties in the social network of *Susceptible* respectively. On the other hand, *Normal Believer Activation*

Fraction and *Average Believer Active Duration* are related to the behavioral responses of people who believe misinformation. Thus, none of these parameters propose a leverage point that can be easily targeted by an intervention to eradicate the spread of false information. Interestingly, none of the changes in *Disbeliever* parameters can prevent the full spread in the current parameter settings although they are somewhat effective in restricting the spread which imply that those parameters might be promising to develop efficient mitigation strategies.

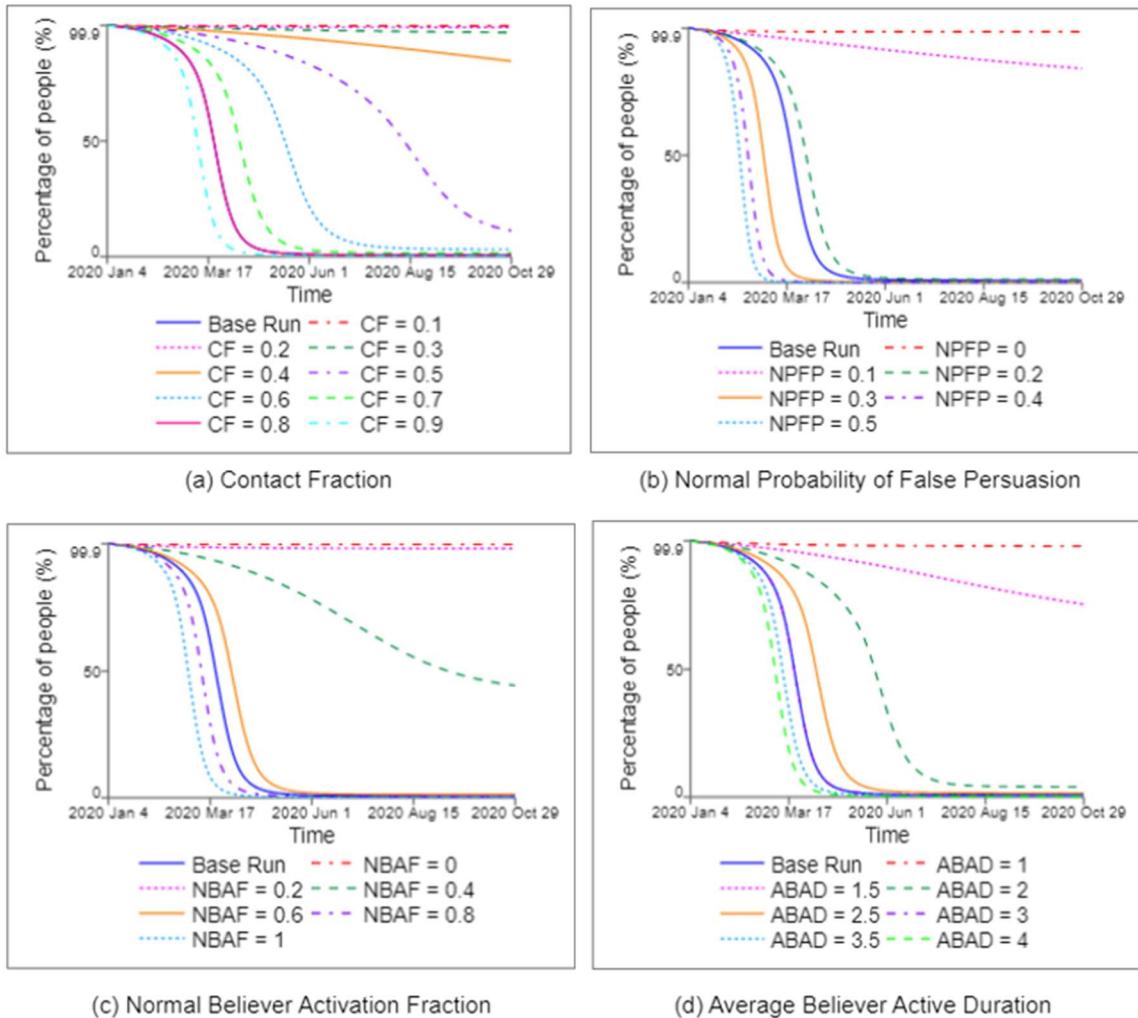


Figure 7.4. Sensitivity of *Susceptible* to different parameters

Another observation is that changes in *Believer* parameters typically create a unidirectional change in the outcomes whereas responses to changes in *Disbeliever* parameters often have nonlinear outcomes. For example, increasing *Believer Activation*

Fraction (Figure C.6) or *Believer Active Duration* (Figure C.12) always results in worse outcomes as observed by increased *Believer Incidence Percentage* and *Total Believer Peak Percentage*. However, increasing *Disbeliever Activation Fraction* (Figure C.9) or *Disbeliever Active Duration* (Figure C.15) produces changed outcomes based on the trade-offs created by Disbeliever Induced Exposure (R2), Disbeliever Induced Persuasion (R3), and Mutual Escalation (R5) which will be further analyzed in the upcoming section.

8. SCENARIO ANALYSIS

Thus far, the model structure and its credibility is discussed and the base behavior of the model has been analyzed. This chapter aims to expand our analysis by investigating what-if scenarios and analyzing changes in the outcome measures for these scenarios. The first two scenarios, “Neutral Sharing” and “Super-spreader” involve slight additions to the model structure which are explained in the corresponding section. The last scenario is obtained by slight changes in the base model parameters or by combining those changes with “Neutral Sharing” and “Super-spreader”.

8.1. Base Run with Neutral Sharing

There is a growing body of evidence that people can share misinformation without essentially considering its veracity. Pennycook and colleagues (2020) found that people’s ability to distinguish between true and false information is significantly low when they are asked whether they would share that information compared to the case when they are asked about the veracity of the information. Thus, sharing behavior does not necessarily require belief in the misinformation but can have many other motives or reasons such as intuitive thinking, self-promotion, and even the thought that misinformation being true would be “interesting” (Ecker et al., 2022). Moreover, for our specific case 5G – COVID 19 conspiracy, there is also supporting evidence that financially motivated agents contribute to the spread (Langguth et al., 2022). Clearly, each of these types of sharing behaviors has different internal drivers and can be incorporated separately into the current model. However, such an investigation is beyond the scope of this work. Instead, we simplify the possible other reasons and just investigate whether the involvement of *Neutral* stock both in the exposure and misinformation generation will create a substantial behavior change in the current model.

To incorporate such a scenario into the model we define *Neutral Engagement Fraction* as the fraction of *Neutral* that generates *Misinformation* and also contributes to its spread. Thus, *Neutral * Neutral Engagement Fraction* represents the number of people

that acts as *Believer Active* even though they do not necessarily believe the misinformation. We denote this amount as *Engaged Neutral*. To generate such a behavior *Engaged Neutral* is added in *Total Active* for exposure and *Misinformation Generation* flow. A simplification made is that engaged neutral people contribute to spreading and generation as much as *Believer Active*.

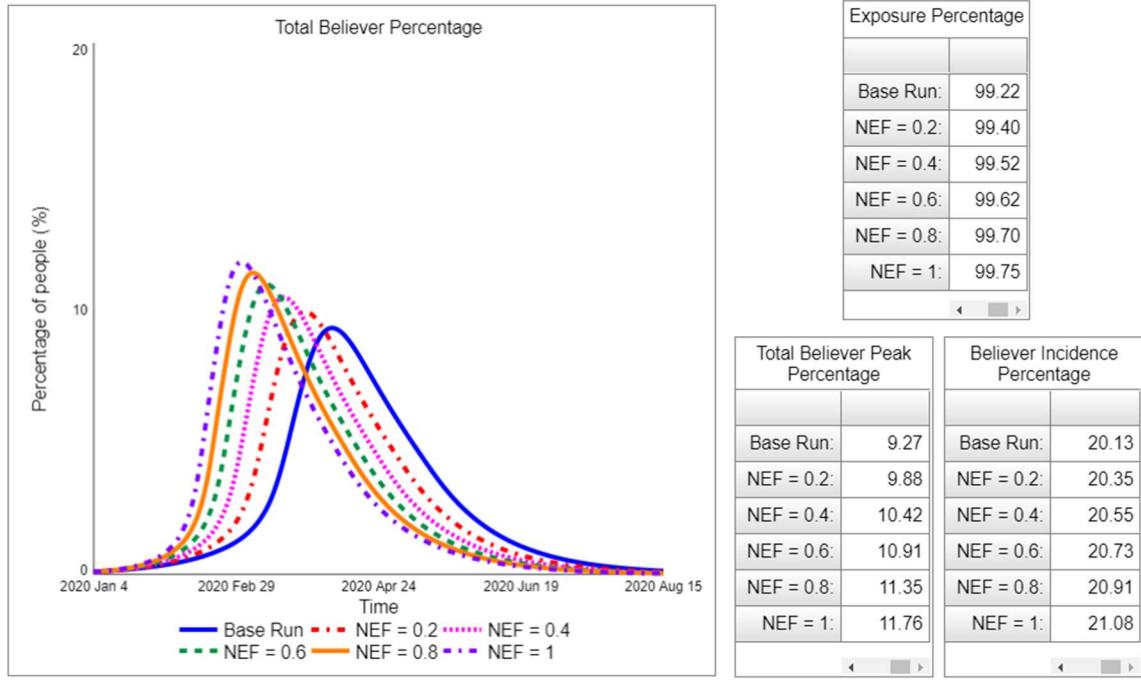


Figure 8.1. The comparative plots of *Total Believer Percentage* and comparative tables of output measures for different values of *Neutral Engagement Fraction*.

Figure 8.1 depicts the comparative plot of *Total Believer* for different values of *Neutral Engagement Fraction*. The main impact of increased *Neutral Engagement Fraction* is the earlier peak observed in the *Total Believers*. Naturally, sharing behavior of neutrals increase *Exposure Rate* as *Susceptible* people can also get exposed by getting in contact with random interactions with neutral people. The effect on *Total Believer Peak Percentage* seems more prominent compared to *Believer Incidence Percentage* as for the former peak percentage has increased from 9.27% to 11.76% whereas for the latter the increase in believer incidence (from 20.13% to 21.08%) is somehow limited.

8.2. Base Run with Super-spreader

Bruns, Harrington, and Hurcombe (2020) have analyzed how the 5G - COVID 19 conspiracy has spread in digital spaces by analyzing Facebook posts. One specific phase of propagation in their analysis includes the spread of conspiracy through celebrities. The amplitude of the exposure through these celebrity accounts reaches enormous numbers on the scale of millions. Rooting from this idea, we wanted to examine the impact of having such a “super-spreader” on the propagation dynamics and compare the measures with the base run.

Parameter Name	Unit	Value
Super-spreader Start Time	day	10 - 130
Super-spreader Contact Fraction	day -1	0.1
Super-spreader Popularity Duration	day	3
Super-spreader Misinformation Generation	information/day	400

Table 8.1. Parameters used for Super-spreader scenarios

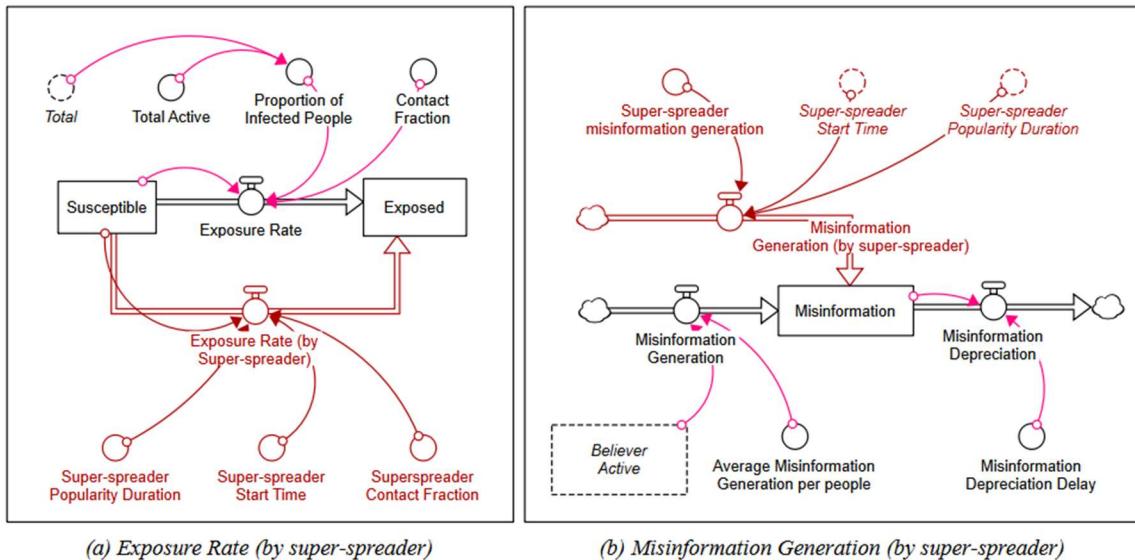


Figure 8.2. Modified Stock-Flow diagrams after incorporating super-spreader. Modified parts are presented in red.

Having a super-spreader sharing such information would have a two-fold effect: a sudden increase in the *Exposed* and *Misinformation* stocks. Thus, the effects are

incorporated into the model via two flows: *Exposure Rate (by Super-spreader)* from *Susceptible* to *Exposed* stock and *Misinformation Generation (by Super-spreader)* from cloud to *Misinformation* stock (Figure 8.2). The two flows become active after *Super-spreader Start Time* and remain active for *Super-spreader Popularity Duration*. For this analysis, we analyzed several values for Super-spreader Start Time and kept other values constant. All of the related parameter values are presented in Table 8.1.

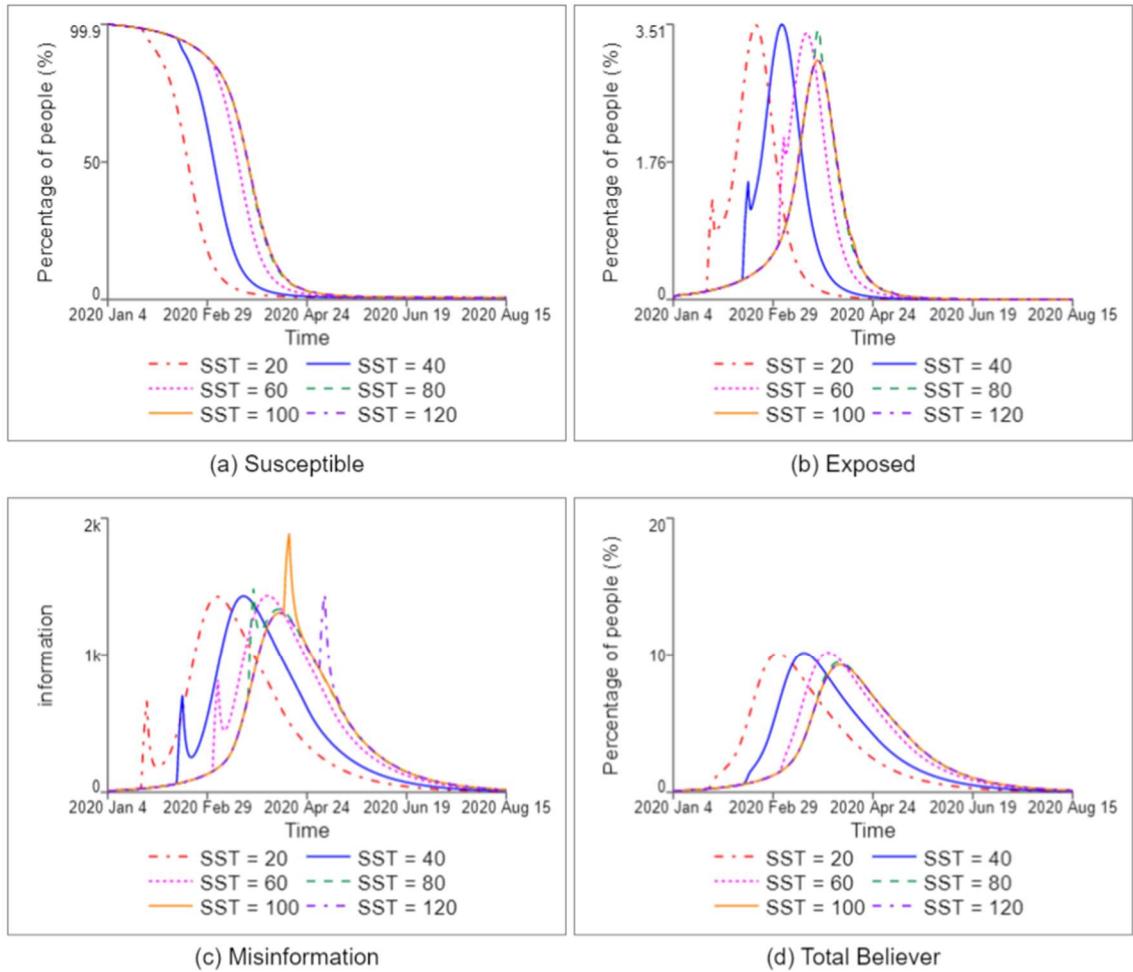


Figure 8.3. Comparative plots for different start times of super-spreader.

The resulting dynamics of four stock variables are presented in Figure 8.3 and the outcome measures for the same run are provided in Table 8.2. The immediate effect of the super-spreader is the mobilization of people from *Susceptible* to *Exposed* (Figure 8.3.a -

8.3.b). When *Super-spreader Start Time* is larger than or equal to 100, there is almost no effect in terms of exposure since the *Susceptible* has been deployed by then. The effect on *Misinformation* is still present after 100 as can be observed by the sharp peaks in Figure 8.3.c. However, since the upstream stocks are already deployed, i.e. misinformation has already gone viral, having a super-spreader does not change the outcome of interests.

Exposure Percentage		Total Believer Peak Percentage		Believer Incidence Percentage	
	Final		Final		Final
SST = 20:	99.40	SST = 20:	10.11	SST = 20:	20.28
SST = 40:	99.41	SST = 40:	10.14	SST = 40:	20.35
SST = 60:	99.39	SST = 60:	10.16	SST = 60:	20.92
SST = 80:	99.26	SST = 80:	9.46	SST = 80:	20.24
SST = 100:	99.25	SST = 100:	9.27	SST = 100:	20.14
SST = 120:	99.25	SST = 120:	9.27	SST = 120:	20.15

Table 8.2. Comparative tables of the outcomes of interest for different start times of superspreaders.

Considering Table 8.2, we can say that the earlier the super-spreader acts, the higher the damage it inflicts in terms of *Total Believer Peak Percentage*. If the super-spreader acts after misinformation have already spread, then the damage is nearly zero since outcomes after start time 100 are nearly the same as their base run values for all three outcomes of interest. Parallel to the observation made for the Neutral Sharing scenario, the adverse effect of having a super-spreader is relatively higher for *Total Believer Peak Percentage* as compared to the *Believer Incidence Percentage*.

8.3. Lower Believability with Neutral Sharing and Super-Spreader

As discussed in Section 7.2 for sufficiently low levels of *Normal Probability of False Persuasion*, we observe almost no spread of misinformation. Thus, this scenario aims to analyze whether the involvement of super-spreaders or neutral sharing can trigger the epidemic in such cases. To analyze such a scenario, we decrease the *Normal Probability of False Persuasion* from 0.22 to 0.1 and generate the base scenario for the “Low

Believability” case. Then, we add super-spreader and normal sharing cases separately. For the *Neutral Engagement Fraction (NEF)*, we experiment with values 0.5 and 0.9. For super-spreader cases, we try four scenarios by changing the *Superspread Start Time (SST)* (50 - 100) and changing the effectiveness of the super-spreader (High, Low).

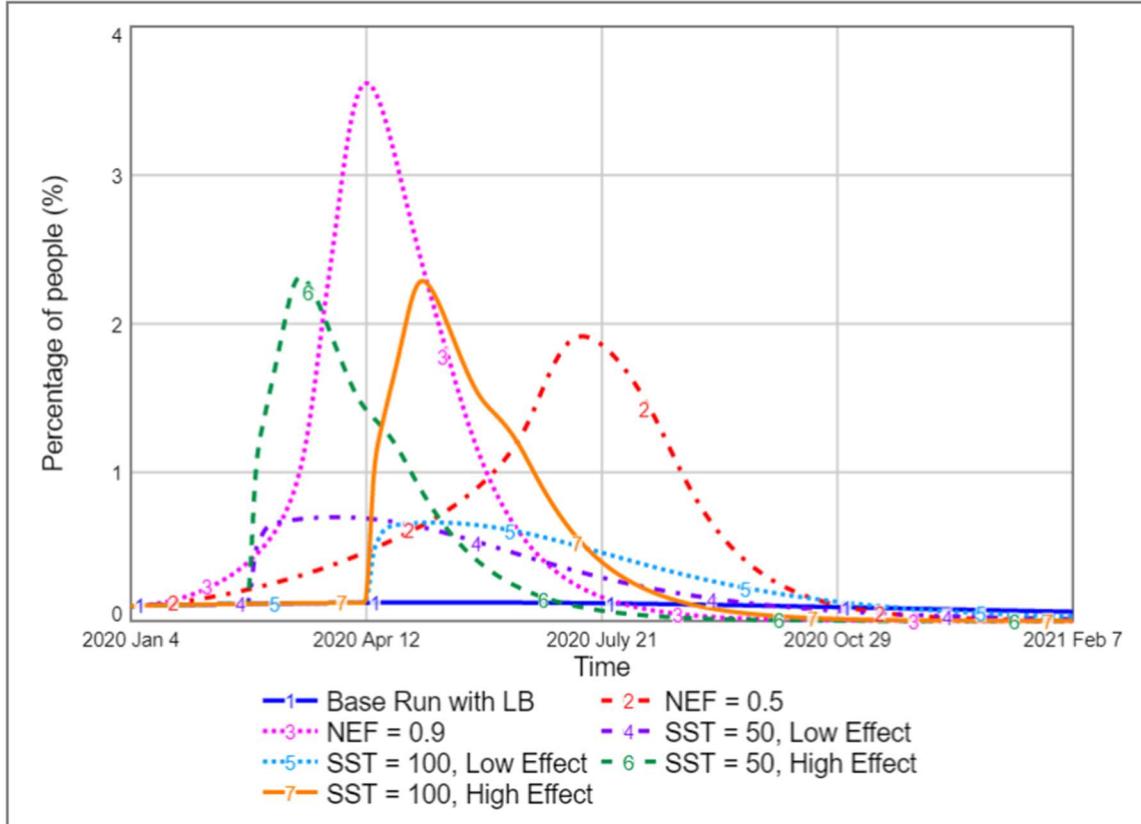


Figure 8.4. Comparative plots of *Total Believer Percentage* dynamics for different scenarios.

The resulting dynamics of *Total Believer* for different cases are presented in Figure 8.4 and the corresponding outcomes of interest tables are presented in Table 8.3. Considering the “Base Run with Low Believability”, the misinformation spread is quite contained with *Believer Incidence Percentage* at 1.75% and *Exposure Rate* is less than 20%. *Total Believer* is nearly constant during the simulation horizon and the *Total Believer Peak Percentage* is 0.12%.

The impact of Neutral Engagement is huge even when the engagement is small as 0.5. Considering Table 8.3, when $NEF = 0.5$, The *Exposure Percentage* jumps from

19.16% to 91.38% compared to the baseline. Therefore, we can say that the involvement of agents other than *Believers* or *Disbelievers* could alter the system behavior and help the epidemic to pass the threshold value. Comparing the two runs with varying *NEF* values we can say that difference in believer incidence and exposure is small as compared to the increase in the peak percentage.

Exposure Percentage		Total Believer Peak Percentage		Believer Incidence Percentage	
	Final		Final		Final
Base Run with LB:	19.16	Base Run with LB:	0.12	Base Run with LB:	1.75
NEF = 0.5:	91.38	NEF = 0.5:	1.91	NEF = 0.5:	8.00
NEF = 0.9:	97.63	NEF = 0.9:	3.62	NEF = 0.9:	8.36
SST = 50, Low Effect:	47.08	SST = 50, Low Effect:	0.70	SST = 50, Low Effect:	4.49
SST = 100, Low Effect:	45.61	SST = 100, Low Effect:	0.66	SST = 100, Low Effect:	4.34
SST = 50, High Effect:	88.85	SST = 50, High Effect:	2.31	SST = 50, High Effect:	5.23
SST = 100, High Effect:	88.70	SST = 100, High Effect:	2.29	SST = 100, High Effect:	5.59

Table 8.3. Comparative tables of the outcomes of interest for different start times of super-spreader.

Regarding super-spreader scenarios, looking at the *Exposure Percentage* (Table 8.3), similar to neutral engagement, super-spreaders can also trigger an infodemic of false information although the damage is limited for the “Low Effect” cases. The timing of the super-spreader seems to create a little impact on the outcomes of interest since the base scenario with low believability almost follows a stabilized spread.

With the current parametrization, the resulting *Total Believer Peak Percentage* values for “High Effect” super-spreaders are as high as the neutral engagement case with $NEF = 0.5$. On the other hand, contrasting the believer incidence values for the same scenarios, neutral engagement produces higher values in comparison to super-spreaders. Moreover, the *Exposure Percentages* for these three runs are quite close. Thus, the lesser believer incidence in super-spreader scenarios is not caused by less exposure, rather sustaining the effect of neutral engagement is more effective in increasing the *Probability of False Persuasion* thus resulting in less incidence. Conversely, the sudden effect of super-spreaders is more impactful in increasing the peak value without affecting the *Believer Incidence Percentage* much.

9. POLICY INTERVENTIONS

9.1. Decreased Disbeliever Activation Fraction

In their work, Ahmed and colleagues (2020) suggest the lesser interaction of the disbeliever group with the believers would be a better option in terms of isolation of the contagion. Thus, one policy can be designed to target the involvement of disbelievers in the debate by either endorsing or discouraging it. To analyze the effectiveness of such a strategy we experiment with different values of *Normal Disbeliever Activation* within the interval $[0, 0.4]$.

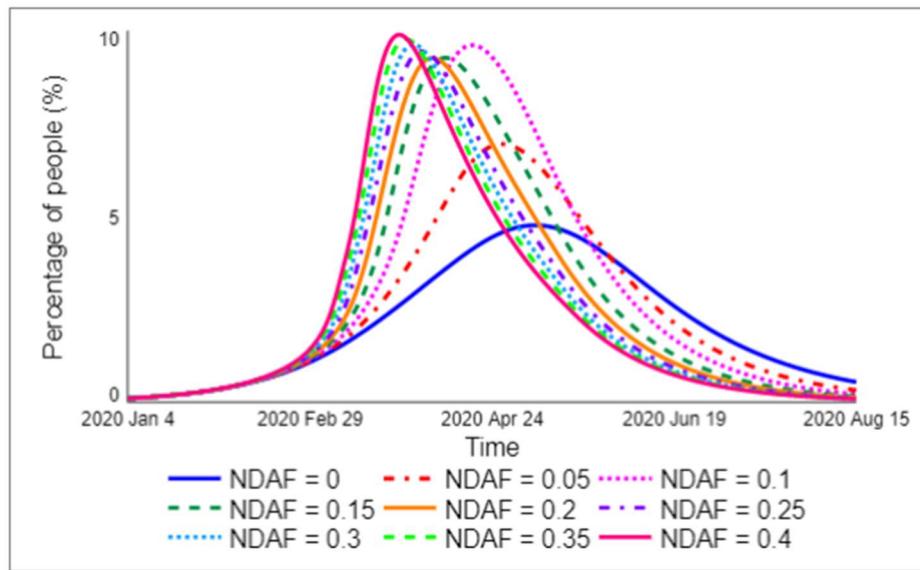


Figure 9.1. The comparative plot of *Total Believer Percentage* for different values of *Normal Disbeliever Activation Fraction (NDAF)*

Simulation results are presented in Figure 9.1 and subsequent measures are presented in Table 9.1. The percentage values in Table 9.1 are colored with respect to the given ‘High’ and ‘Low’ values in the lower part of the table. Considering the *Total Believer Peak Percentage* the optimal outcome is obtained when the activation of disbelievers is zero i.e. when none of the disbelievers share their opinion or generate corrective information on the subject. Thus, decreasing the *Disbeliever Activation* seems favorable intervention while trying to minimize the peak value of believers. On the other hand, increasing activation further seems to have little to no effect on *Total Believer Peak*

Percentage although the peak value shows a small increase with the increasing disbeliever activation. *Exposure Percentage* follows the same pattern as the peak value, since the minimum exposure is obtained for the case of no activation of disbelievers and the maximum is obtained when *Normal Disbeliever Activation Fraction* is at maximum.

	Total Believer Peak Percentage	Believer Incidence Percentage	Exposure Percentage
NDAF (0):	4.8	20.7	91.4
NDAF (0.05):	7.0	23.2	95.8
NDAF (0.1):	9.6	23.8	98.6
NDAF (0.15):	9.3	21.5	98.9
NDAF (0.2):	9.3	20.3	99.2
NDAF (0.25):	9.5	19.6	99.5
NDAF (0.3):	9.6	19.2	99.7
NDAF (0.35):	9.8	18.8	99.8
NDAF (0.4):	9.9	18.5	99.9
Low	5.0	15.0	90.0
Base Run	9.3	20.3	99.2
High	15.0	25.0	100.0

Table 9.1. Comparative table of three output measures for different values of *Normal Disbeliever Activation Fraction (NDAF)*

As opposed to the other two measures *Believer Incidence Percentage* does not improve as the *Normal Disbeliever Activation Fraction* decreases. Furthermore, the outcome is far away from being linear. Starting from zero, increasing *Normal Disbeliever Activation Fraction (NDAF)* results in a higher *Believer Incidence Percentage* initially, whereas after 0.1 we observe better outcomes with the best value of 18.5 % is achieved when the *Normal Disbeliever Activation Fraction* is 0.4.

Considering Table 9.1, it is evident that optimal outcomes for two measures, *Total Believer Peak Percentage* and *Believer Incidence Percentage*, are obtained in different parameter settings. Furthermore, the strategy to improve the Base Run ($NDAF = 0.2$), either by promoting or discouraging the activation of disbelievers, also depends on the selected outcome of interest.

To understand the multifold effects of such intervention, comparative plots of four stocks are presented in Figure 9.2. To simplify the representation, only 5 of the runs are presented.

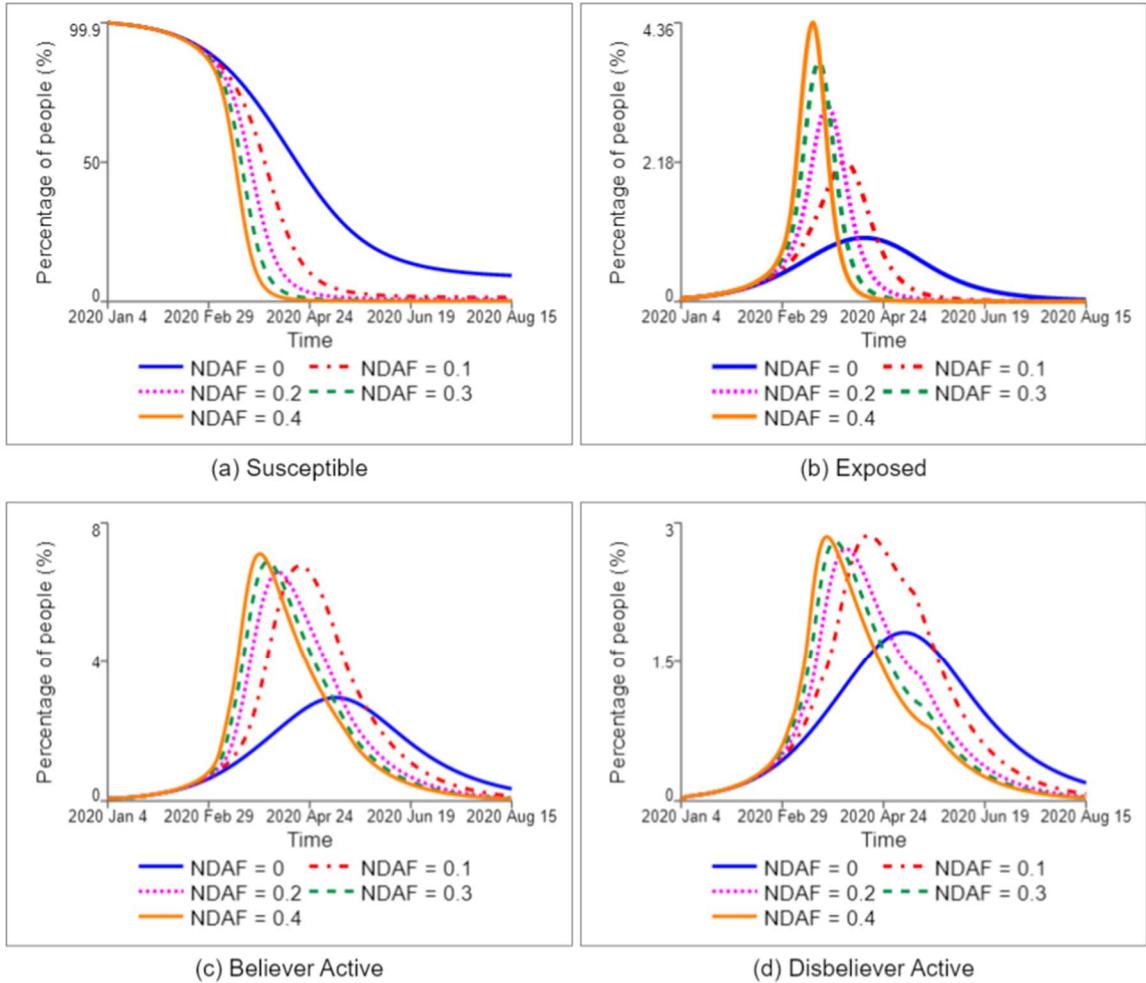


Figure 9.2. Comparative plots of dynamics of four stocks for different values of *Normal Disbeliever Activation Fraction (NDAF)*

One direct effect of increased *Normal Disbeliever Activation Fraction* is the increased *Exposure Percentage* (Table 9.1). The susceptible people become exposed to misinformation as they get in contact with either *Believer Active* or *Disbeliever Active* people. Therefore, if *Normal Disbeliever Activation Fraction* is constrained, it results in less *Disbeliever Active* (Figure 9.2.d) and a smaller number of people getting exposed to misinformation. Conversely, if *Normal Disbeliever Activation Fraction* is increased, a

higher number of *Disbeliever Active* would directly increase the flow from *Susceptible* to *Exposed* (Figure 9.2.a, 9.2.b).

An indirect but parallel effect is observed through the Mutual Escalation Loop (R5). As mentioned earlier, as the number of people actively contributing to the debate increases, it also encourages more people from the opposing side to become active. Therefore, increased *Normal Disbeliever Activation Fraction* also increases the *Believer Active* (Figure 9.2.c) which again directly increases the number of people mobilized from *Susceptible* to *Exposed* (Figure 9.2.a, 9.2.b).

Considering only the *Exposure Percentage*, these two pathways synergically work which is evident in a parallel increase in *Exposure Percentage* in response to an increase in *Normal Disbeliever Activation Fraction* (Table 9.1). The simplicity in this relation is rooted in the fact that *Exposure Percentage* is not affected by the information dominance in the digital sphere rather it is only a measure of the virality of the misinformation independent of whether it is believed or disbelieved.

Same direct (increase in *Disbeliever Active*) and indirect (increase in *Believer Active* through Mutual Escalation (R5)) pathways, however, become antagonists when we consider the effect of changing *Normal Disbeliever Activation Fraction* on *Actual Probability of False Persuasion*. Considering the direct pathway, the more disbelievers involved in the discussion the more *Corrective Info* is generated by disbelievers. Conversely, through Mutual Escalation Loop (R5), increased *Believer Active* will result in increased levels of *Misinformation*. Thus, two information types compete in their effect on *Probability of False Persuasion*. The resulting *Actual Probability of False Persuasion* is presented in Figure 9.3. It can be seen that for the smaller values of *NDAF* such as 0.1 the Actual Probability of False Persuasion is greater than no activation case. Thus, for the small activation case, the effect through Mutual Escalation Loop (R5) dominates the direct increase in *Disbeliever Active*. However, when the *NDAF* is high enough the increase in *Disbeliever Active* dominates its indirect effect on *Believer Active* resulting in less probability of believing misinformation.

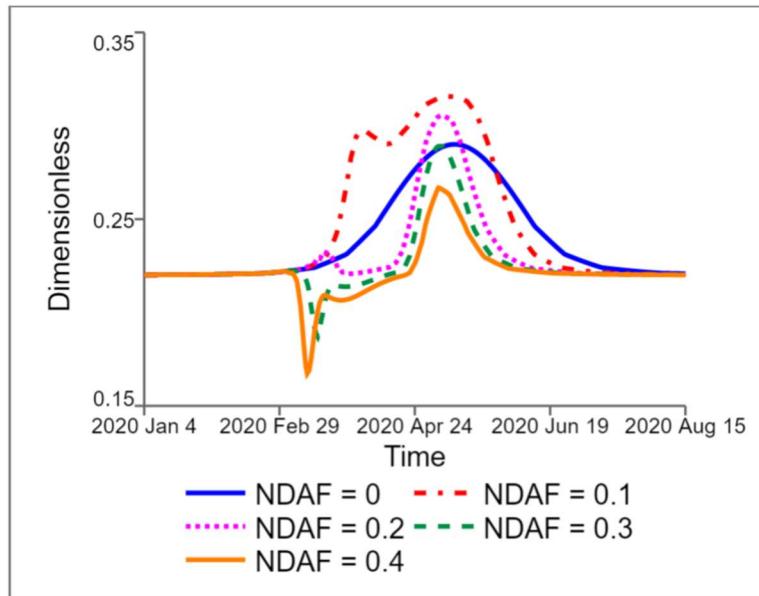


Figure 9.3. The comparative plots of *Actual Probability of False Persuasion* for different values of *Normal Disbeliever Activation Fraction*

9.2. Corrective Information Campaign

One of the main criticisms of the management of the 5G-COVID 19 conspiracy spread was the lack of explanations and denouncements by the authority figures (Ahmed et al., 2020). Using this idea, an “Information Campaign” on the social media platform is tested as a mitigation strategy. The information campaign is implemented as an inflow to the *Corrective Information* starting at *Campaign Start* for the duration of *Campaign Duration*. The amount of inflow is denoted as *Campaign Intensity* and assumed to be constant during the campaign. The analysis is conducted for different values of *Campaign Start* and *Campaign Duration*, whereas *Campaign Intensity* is kept constant (equal to 300) for the analysis to simplify the comparisons.

The complete analysis of all three of the outcomes of interest for different scenarios is presented in Appendix D. Since the insights derived are mostly consistent for all outcomes of interest, only results of *Believer Incidence Percentage* are discussed in this section. Table 9.2 shows the final *Believer Incidence Percentages* for different values of

Campaign Start and *Campaign Duration*. The color scale is applied based on the “Base Run” value 20.25% for white and “lowest” value 15.51% for green.

Believer Incidence		Campaign Duration				
	Percentage	10	20	30	40	50
Campaign Start (date)	40	19.95	19.41	18.26	15.99	15.51
	50	19.6	17.69	16.18	15.97	15.84
	60	18.1	16.97	16.76	16.65	16.56
	70	18.49	18.28	18.17	18.09	18
	80	19.98	19.87	19.8	19.72	19.68
	90	20.15	20.07	20	19.96	19.94
	100	20.18	20.1	20.06	20.04	20.03
	110	20.15	20.11	20.09	20.08	20.08
	120	20.2	20.18	20.17	20.17	20.16

Table 9.2. Comparative table of *Believer Incidence Percentage* for different values of *Campaign Start* and *Campaign Duration*. The color scale corresponds to “base run” and “low” values for white and green respectively.

Given a fixed start time, increasing the duration of the campaign seems to be effective for all starting dates which presents a consistent scenario with intuitive thinking. However, although the change of direction remains the same, the observed improvement is very little for the interventions after 70, given that the *Believer Incidence Percentage* was 20.25 % for the base case scenario without the information campaign

Regarding the start time of the campaign, unidirectional thinking would suggest intervening as early as possible would produce better outcomes. However, results indicate that early interventions might be ineffective if the duration of the intervention is not lasting. Given that the misinformation peaks between days 80 – 90, we can deduce that for such a campaign to be used at maximum effectiveness, it should be sustained until the peak in total believers is achieved.

10. CONCLUSIONS

We have reviewed the literature on misinformation spread on social media specifically for the 5G-COVID-19 narrative and built a system dynamics model for the problem. The model is constructed using both quantitative and qualitative literature and validation with respect to real data and extreme condition tests are presented. Scenario analysis is presented for super-spreader, neutral engagement, and low believability cases. In addition, using the proposed interventions from the literature, policy analysis is conducted for two different interventions, namely decreasing *Disbeliever Activation Fraction* and intervention of corrective information campaign.

Scenario analysis presents that for misinformation that would go viral without neutral engagement or super-spreaders, the addition of such cases does not create considerable differences in the outcome of interests. However, the experiment with misinformation with lower believability revealed that incidence of such scenarios might generate substantial changes by triggering the reinforcing loops and resulting in an epidemic. Super-spreaders constitute a higher risk in terms of increasing peak percentage whereas they are somewhat less effective in believer incidence as compared to neutral sharing.

The results from the policy analysis indicate that the nonlinear relationships in the system present several trade-offs. Firstly, the outcomes of interest do not always follow the same direction of change for the policy interventions. As an example, regarding the *Believer Incidence Percentage*, increasing *Disbeliever Activation Fraction* unidirectionally produces better outcomes (Table 9.1). On the other hand, the same policy would produce worse outcomes in terms of *Total Believer Peak Percentage* as evident in the increased level of peak values in Figure 9.1. Such a difference is important as one outcome might be more important for specific types of misinformation. For instance, in 5G-COVID 19 case, looking at the reported violence in the comments on social media platforms even in the early stages (Bruns, Harrington, and Hurcombe, 2020) policymakers can focus on minimizing the maximum number of believers to avoid violent protests but, for a case about

information that can change long-term behaviors such as climate action, decision-makers may prioritize minimizing the total number of believers during the spread.

Another observation is that decreasing *Disbeliever Activation* which is commonly cited as a prevention method in the literature does not follow a unidirectional pattern in terms of effectiveness. Our analysis indicates that limiting the involvement of the disbelievers in the online discussion does not necessarily produce better outcomes for *Believer Incidence* but rather presents a trade-off between exposing more people or generating more corrective information to inform people. The model at hand is calibrated for a specific case thus generalizability of numerical results to other cases can be questionable. However, the model is built upon one of the simplest information diffusion models (SIR) with rather small changes in the model structure. Despite such simplicity and very few causal effects included, it still presents a trade-off that indicates the complexity of the problem at hand. In addition, the causally descriptive nature of the System Dynamics model allowed us to analyze the reasons for the observed behavior which paves the way for extending the analysis by modifying and calibrating the model for other cases of misinformation. Such extension would also provide structural analogies between different cases of misinformation which is a promising path for future work.

Results from the policy analysis of the information campaign indicate that such an intervention should be planned carefully to maximize the improvements in the outcomes. If the duration of the campaign is limited then it should be timed carefully, while if such a limitation is not present the impact will be maximum when the intervention starts as early as possible and is sustained until the peak value of the exposed people is achieved.

To sum, the analysis of the model represents several trade-offs that can result in unintended consequences for the proposed mitigation strategies in the literature and the designed scenario analysis indicates the inclusion of neutral sharing or super-spreaders can create substantial change in the behavior. The future research for this research is to test the model assumptions, outputs, and robustness of the insights further by making use of richer cross-sectional and dynamic data. Moreover, other possible mitigation strategies can be incorporated into the model to assess the effectiveness for different scenarios. Another

agenda is to expand the model by including more user profiles such as “like-seekers” instead of just two opposing sides as there are different motivations for other groups to engage. Different susceptibility of these groups may also be incorporated as such a classification and differentiation is presented in the current literature (Agley and Xiao, 2021). Finally, a discussion on similarities and disparities between our specific case and other types of misinformation spread can be useful to infer the harmony of interventions for different cases of false information spread.

APPENDIX A : MODEL EQUATIONS

Top-Level Model:

$$\text{Believer_Active}(t) = \text{Believer_Active}(t - dt) + (\text{Believer_Activation_Rate} - \text{Believer_Deactivation_Rate}) * dt \quad \{\text{NON-NEGATIVE}\}$$

$$\text{INIT Believer_Active} = \text{Active_Initial}$$

UNITS: People

$$\text{Believer_Dormant}(t) = \text{Believer_Dormant}(t - dt) + (\text{Believer_Deactivation_Rate} + \text{Believer_Adoption_Rate} - \text{Believer_Activation_Rate} - \text{Believer_Quit_Rate}) * dt \quad \{\text{NON-NEGATIVE}\}$$

$$\text{INIT Believer_Dormant} = 0$$

UNITS: People

$$\text{Corrective_Info}(t) = \text{Corrective_Info}(t - dt) + (\text{Corrective_Info_Generation} + \text{Info_Campaign} - \text{Corrective_Info_Depreciation}) * dt \quad \{\text{NON-NEGATIVE}\}$$

$$\text{INIT Corrective_Info} = 0$$

UNITS: information

$$\text{Cumulative_Believer_Tweets}(t) = \text{Cumulative_Believer_Tweets}(t - dt) + (\text{Posted_Believer_Tweets}) * dt \quad \{\text{NON-NEGATIVE}\}$$

$$\text{INIT Cumulative_Believer_Tweets} = 0$$

UNITS: Tweets

$$\text{Disbeliever_Active}(t) = \text{Disbeliever_Active}(t - dt) + (\text{Disbeliever_Activation_Rate} - \text{Disbeliever_Deactivation_Rate}) * dt \quad \{\text{NON-NEGATIVE}\}$$

$$\text{INIT Disbeliever_Active} = 0$$

UNITS: People

$$\text{Disbeliever_Dormant}(t) = \text{Disbeliever_Dormant}(t - dt) + (\text{Disbeliever_Deactivation_Rate} + \text{Disbeliever_Adoption_Rate} - \text{Disbeliever_Activation_Rate} - \text{Disbeliever_Quit_Rate}) * dt \quad \{\text{NON-NEGATIVE}\}$$

$$\text{INIT Disbeliever_Dormant} = 0$$

UNITS: People

$$\text{Exposed}(t) = \text{Exposed}(t - dt) + (\text{Exposure_Rate} + \text{"Exposure_Rate_by_Super-spreaders"}) - \text{Disbeliever_Adoption_Rate} - \text{Believer_Adoption_Rate} - \text{Neutral_Adoption_Rate} * dt \quad \{\text{NON-NEGATIVE}\}$$

INIT Exposed = 0

UNITS: People

$$\text{Misinformation}(t) = \text{Misinformation}(t - dt) + (\text{Misinformation_Generation} + \text{"Misinformation_Generation_by_super-spreader"}) - \text{Misinformation_Depreciation} * dt \quad \{\text{NON-NEGATIVE}\}$$

INIT Misinformation = 0

UNITS: information

$$\text{Neutral}(t) = \text{Neutral}(t - dt) + (\text{Neutral_Adoption_Rate} - \text{Neutral_Quit_Rate}) * dt \quad \{\text{NON-NEGATIVE}\}$$

INIT Neutral = 0

UNITS: People

$$\text{Susceptible}(t) = \text{Susceptible}(t - dt) + (-\text{Exposure_Rate} - \text{"Exposure_Rate_by_Super-spreaders"}) * dt \quad \{\text{NON-NEGATIVE}\}$$

INIT Susceptible = S_Initial

UNITS: People

$$\text{Total_Active_AUC}(t) = \text{Total_Active_AUC}(t - dt) + (\text{Total_Active_AUC_increase}) * dt \quad \{\text{NON-NEGATIVE}\}$$

INIT Total_Active_AUC = 0

UNITS: People*days

$$\text{Total_Believer_AUC}(t) = \text{Total_Believer_AUC}(t - dt) + (\text{Total_Believer_AUC_increase}) * dt \quad \{\text{NON-NEGATIVE}\}$$

INIT Total_Believer_AUC = 0

UNITS: People*days

Total_Quit_from_Believer(t) = Total_Quit_from_Believer(t - dt) +
(Total_Quit_from_Believer_Increase) * dt {NON-NEGATIVE}

INIT Total_Quit_from_Believer = 0

UNITS: People

Total_Quit_from_Disbeliever(t) = Total_Quit_from_Disbeliever(t - dt) +
(Total_Quit_from_Disbeliever_Increase) * dt {NON-NEGATIVE}

INIT Total_Quit_from_Disbeliever = 0

UNITS: People

Total_Quit_from_Neutral(t) = Total_Quit_from_Neutral(t - dt) +
(Total_Quit_from_Neutral_Increase) * dt {NON-NEGATIVE}

INIT Total_Quit_from_Neutral = 0

UNITS: People

Believer_Activation_Rate = Believer_Dormant * Actual_Believer_Activation_Fraction
{UNIFLOW}

OUTFLOW PRIORITY: 1

UNITS: person/days

Believer_Adoption_Rate = Exposed * (1 - Neutral_Fract) * Actual_Probability_of_False_Persuasion / Believer_Adoption_Time
{UNIFLOW}

OUTFLOW PRIORITY: 2

UNITS: person/days

Believer_Deactivation_Rate = Believer_Active / Average_Believer_Active_Duration
{UNIFLOW}

UNITS: person/days

Believer_Quit_Rate = Believer_Dormant / Believer_Quit_Time {UNIFLOW}

OUTFLOW PRIORITY: 2

UNITS: People/days

Corrective_Info_Depreciation = Corrective_Info/Corrective_info_Depreciation_Time
{UNIFLOW}

UNITS: information/days

Corrective_Info_Generation =
Disbeliever_Active*Average_Corrective_Info_Generation_Per_people {UNIFLOW}

UNITS: information/days

Disbeliever_Activation_Rate =
Disbeliever_Dormant*Actual_Disbeliever_Activation_Fraction {UNIFLOW}

OUTFLOW PRIORITY: 1

UNITS: person/days

Disbeliever_Adoption_Rate = Exposed*(1-Neutral_Fract)*(1-
Actual_Probability_of_False_Persuasion)/Disbeliever_Adoption_Time {UNIFLOW}

OUTFLOW PRIORITY: 1

UNITS: person/days

Disbeliever_Deactivation_Rate =
Disbeliever_Active/Average_Disbeliever_Active_Duration {UNIFLOW}

UNITS: person/days

Disbeliever_Quit_Rate = Disbeliever_Dormant/Disbeliever_Quit_Time {UNIFLOW}

OUTFLOW PRIORITY: 2

UNITS: People/days

Exposure_Rate = Susceptible*(Contact_Fraction*Proportion_of_Infected_People)
{UNIFLOW}

OUTFLOW PRIORITY: 1

UNITS: People/days

"Exposure_Rate_(by_Super-spreaders)" = (STEP(1, "Super-spreader_start_time")-
STEP(1, "Super-spreader_start_time"+"Super-
spreader_popularity_duration"))*Susceptible*Superspreaders_Contact_fraction*"IsSupers
preader_(bool)" {UNIFLOW}

OUTFLOW PRIORITY: 2

UNITS: People/days

Info_Campaign = STEP(1, Campaign_Start)*Campaign_Intensity - STEP(1, Campaign_Start+Campaign_Duration)*Campaign_Intensity {UNIFLOW}

UNITS: information/days

Misinformation_Depreciation = Misinformation/Misinformation_Depreciation_Time {UNIFLOW}

UNITS: information/days

Misinformation_Generation =

Believer_Active*Average_Misinformation_Generation_per_people+Neutral*Neutral_Misinformation_Generation_Per_people*Neutral_Engagement_Fraction {UNIFLOW}

UNITS: information/days

"Misinformation_Generation_(by_super-spreader)" = (STEP(1, "Super-spreader_start_time")-STEP(1, "Super-spreader_start_time"+"Super-spreader_popularity_duration"))*"Super-spreader_misinformation_generation""IsSuperspreader_(bool)" {UNIFLOW}

UNITS: information/days

Neutral_Adoption_Rate = Exposed*(Neutral_Fract)/Neutral_Adoption_Time {UNIFLOW}

OUTFLOW PRIORITY: 3

UNITS: person/day

Neutral_Quit_Rate = Neutral/Neutral_Quit_Time {UNIFLOW}

UNITS: People/days

Posted_Believer_Tweets = Believer_Active/Average_Time_To_Tweet {UNIFLOW}

UNITS: Tweets/day

Total_Active_AUC_increase = Total_Active {UNIFLOW}

UNITS: People

Total_Believer_AUC_increase = Total_Believer {UNIFLOW}

UNITS: People

Total_Quit_from_Believer_Increase = Believer_Quit_Rate {UNIFLOW}

UNITS: People/days

Total_Quit_from_Disbeliever_Increase = Disbeliever_Quit_Rate {UNIFLOW}

UNITS: People/days

Total_Quit_from_Neutral_Increase = Neutral_Quit_Rate {UNIFLOW}

UNITS: People/days

Active_Initial = 10

UNITS: People

Actual_Believer_Activation_Fraction =

Normal_Believer_Activation_Fraction*Effect_of_Corrective_Info_on_Believer_Activation_Fraction

UNITS: Per Day

Actual_Disbeliever_Activation_Fraction =

Normal_Disbeliever_Activation_Fraction*Effect_of_Misinformation_on_Disbeliever_Activation_Fraction

UNITS: Per Day

Actual_Probability_of_False_Persuasion =

Normal_Prob_of_False_Persuasion+Prob_of_False_Persuasion_Effect_Multiplier*Effect_of_Misinformation_on_Prob_of_False_Persuasion+
Effect_of_Corrective_Info_on_Prob_of_False_Persuasion*
Prob_of_False_Persuasion_Effect_Multiplier

UNITS: Dimensionless

Average_Believer_Active_Duration = 3

UNITS: Days

Average_Corrective_Info_Generation_Per_people = 1

UNITS: Information/(Day*People)

Average_Disbeliever_Active_Duration = 1

UNITS: Days

Average_Misinformation_Generation_per_people = 1

UNITS: Information/(Day*People)

Average_Time_To_Tweet = 7

UNITS: People*days/Tweets

Believer_Active_Percentage = 100*Believer_Active/Total

UNITS: fraction

Believer_Adoption_Time = 1

UNITS: Days

Believer_Dormant_Percentage = 100*Believer_Dormant/Total

UNITS: fraction

Believer_Incidence_Percentage = 100*Total_Quit_from_Believer/Total

UNITS: fraction

Believer_Quit_Time = 1/0.11

UNITS: Days

Campaign_Duration = 20

UNITS: Days

Campaign_Intensity = 0

UNITS: information/days

Campaign_Start = 70

UNITS: Day

Contact_Fraction = 0.8

UNITS: Per Day

Corrective_info_Depreciation_Time = 2

UNITS: Days

Corrective_info_per_capita = Corrective_Info/Total

UNITS: information/people

Cumulative_Incidence_Data = GRAPH(TIME)

Points: (0.0, 0), (1.0, 0), (2.0, 0), (3.0, 0), (4.0, 0), (5.0, 0), (6.0, 0), (7.0, 0), (8.0, 0), (9.0, 0), (10.0, 0), (11.0, 0), (12.0, 0), (13.0, 0), (14.0, 0), (15.0, 0), (16.0, 0), (17.0, 0), (18.0, 0), (19.0, 0), (20.0, 0), (21.0, 0), (22.0, 0), (23.0, 131), (24.0, 136), (25.0, 155), (26.0, 179), (27.0, 209), (28.0, 253), (29.0, 262), (30.0, 267), (31.0, 275), (32.0, 279), (33.0, 281), (34.0, 282), (35.0, 298), (36.0, 298), (37.0, 300), (38.0, 301), (39.0, 301), (40.0, 301), (41.0, 301), (42.0, 302), (43.0, 302), (44.0, 305), (45.0, 305), (46.0, 309), (47.0, 310), (48.0, 312), (49.0, 314), (50.0, 314), (51.0, 314), (52.0, 315), (53.0, 315), (54.0, 317), (55.0, 318), (56.0, 321), (57.0, 321), (58.0, 321), (59.0, 324), (60.0, 335), (61.0, 343), (62.0, 347), (63.0, 349), (64.0, 349), (65.0, 353), (66.0, 354), (67.0, 358), (68.0, 359), (69.0, 359), (70.0, 363), (71.0, 365), (72.0, 365), (73.0, 372), (74.0, 375), (75.0, 375), (76.0, 382), (77.0, 404), (78.0, 417), (79.0, 444), (80.0, 464), (81.0, 501), (82.0, 517), (83.0, 542), (84.0, 570), (85.0, 599), (86.0, 625), (87.0, 671), (88.0, 735), (89.0, 800), (90.0, 1023), (91.0, 1333), (92.0, 1592), (93.0, 1911), (94.0, 2113), (95.0, 2453), (96.0, 2592), (97.0, 2709), (98.0, 2774), (99.0, 2888), (100.0, 3017), (101.0, 3163), (102.0, 3257), (103.0, 3352), (104.0, 3475), (105.0, 3533), (106.0, 3597), (107.0, 3675), (108.0, 3740), (109.0, 3780), (110.0, 3822), (111.0, 3925), (112.0, 3973), (113.0, 4031), (114.0, 4068), (115.0, 4100), (116.0, 4172), (117.0, 4206), (118.0, 4229), (119.0, 4277), (120.0, 4380), (121.0, 4444), (122.0, 4507), (123.0, 4559), (124.0, 4590), (125.0, 4621), (126.0, 4642), (127.0, 4661), (128.0, 4686), (129.0, 4703), (130.0, 4710), (131.0, 4728), (132.0, 4751), (133.0, 4767), (134.0, 4786), (135.0, 4791), (136.0, 4805), (137.0, 4813), (138.0, 4827), (139.0, 4834), (140.0, 4849), (141.0, 4881), (142.0, 4912), (143.0, 4918), (144.0, 4922), (145.0, 4941), (146.0, 4944), (147.0, 4951), (148.0, 4951), (149.0, 4953), (150.0, 4957), (151.0, 4963), (152.0, 4966), (153.0, 4967), (154.0, 5049), (155.0, 5379), (156.0, 5388), (157.0, 5407), (158.0, 5412), (159.0, 5416), (160.0, 5423), (161.0, 5427), (162.0, 5428), (163.0, 5429), (164.0, 5435), (165.0, 5445), (166.0, 5448), (167.0, 5456), (168.0, 5460), (169.0, 5461), (170.0, 5462), (171.0, 5464), (172.0, 5467), (173.0, 5467), (174.0, 5473), (175.0, 5474), (176.0, 5479), (177.0, 5481), (178.0, 5481), (179.0, 5483), (180.0, 5485), (181.0, 5487), (182.0, 5489), (183.0, 5494), (184.0, 5498), (185.0, 5499), (186.0, 5503), (187.0, 5507), (188.0, 5513), (189.0, 5514), (190.0, 5519), (191.0, 5519), (192.0, 5523), (193.0, 5524), (194.0, 5525), (195.0, 5527), (196.0, 5529), (197.0, 5530), (198.0, 5531), (199.0, 5533), (200.0, 5538), (201.0, 5551), (202.0, 5555), (203.0, 5557), (204.0, 5562), (205.0, 5586), (206.0, 5587), (207.0, 5588), (208.0, 5589), (209.0, 5590), (210.0, 5592), (211.0, 5595), (212.0, 5598), (213.0, 5600), (214.0, 5600), (215.0, 5600), (216.0, 5600), (217.0, 5600), (218.0, 5605), (219.0, 5607), (220.0, 5608), (221.0, 5609), (222.0, 5610), (223.0, 5610), (224.0, 5611)

UNITS: Tweets

Daily_HashTag_Data = GRAPH(TIME)

Points: (0.0, 0.0), (1.0, 0.0), (2.0, 0.0), (3.0, 0.0), (4.0, 0.0), (5.0, 0.0), (6.0, 0.0), (7.0, 0.0), (8.0, 0.0), (9.0, 0.0), (10.0, 0.0), (11.0, 0.0), (12.0, 0.0), (13.0, 0.0), (14.0, 0.0), (15.0, 0.0), (16.0, 0.0), (17.0, 0.0), (18.0, 0.0), (19.0, 0.0), (20.0, 0.0), (21.0, 0.0), (22.0, 0.0), (23.0, 131.0), (24.0, 5.0), (25.0, 19.0), (26.0, 24.0), (27.0, 30.0), (28.0, 44.0), (29.0, 9.0), (30.0, 5.0), (31.0, 8.0), (32.0, 4.0), (33.0, 2.0), (34.0, 1.0), (35.0, 16.0), (36.0, 0.0), (37.0, 2.0), (38.0, 1.0), (39.0, 0.0), (40.0, 0.0), (41.0, 0.0), (42.0, 1.0), (43.0, 0.0), (44.0, 3.0), (45.0, 0.0), (46.0, 4.0), (47.0, 1.0), (48.0, 2.0), (49.0, 2.0), (50.0, 0.0), (51.0, 0.0), (52.0, 1.0), (53.0, 0.0), (54.0, 2.0), (55.0, 1.0), (56.0, 3.0), (57.0, 0.0), (58.0, 0.0), (59.0, 3.0), (60.0, 11.0), (61.0, 8.0), (62.0, 4.0), (63.0, 2.0), (64.0, 0.0), (65.0, 4.0), (66.0, 1.0), (67.0, 4.0), (68.0, 1.0), (69.0, 0.0), (70.0, 4.0), (71.0, 2.0), (72.0, 0.0), (73.0, 7.0), (74.0, 3.0), (75.0, 0.0), (76.0, 7.0), (77.0, 22.0), (78.0, 13.0), (79.0, 27.0), (80.0, 20.0), (81.0, 37.0), (82.0, 16.0), (83.0, 25.0), (84.0, 28.0), (85.0, 29.0), (86.0, 26.0), (87.0, 46.0), (88.0, 64.0), (89.0, 65.0), (90.0, 223.0), (91.0, 310.0), (92.0, 259.0), (93.0, 319.0), (94.0, 202.0), (95.0, 340.0), (96.0, 139.0), (97.0, 117.0), (98.0, 65.0), (99.0, 114.0), (100.0, 129.0), (101.0, 146.0), (102.0, 94.0), (103.0, 95.0), (104.0, 123.0), (105.0, 58.0), (106.0, 64.0), (107.0, 78.0), (108.0, 65.0), (109.0, 40.0), (110.0, 42.0), (111.0, 103.0), (112.0, 48.0), (113.0, 58.0), (114.0, 37.0), (115.0, 32.0), (116.0, 72.0), (117.0, 34.0), (118.0, 23.0), (119.0, 48.0), (120.0, 103.0), (121.0, 64.0), (122.0, 63.0), (123.0, 52.0), (124.0, 31.0), (125.0, 31.0), (126.0, 21.0), (127.0, 19.0), (128.0, 25.0), (129.0, 17.0), (130.0, 7.0), (131.0, 18.0), (132.0, 23.0), (133.0, 16.0), (134.0, 19.0), (135.0, 5.0), (136.0, 14.0), (137.0, 8.0), (138.0, 14.0), (139.0, 7.0), (140.0, 15.0), (141.0, 32.0), (142.0, 31.0), (143.0, 6.0), (144.0, 4.0), (145.0, 19.0), (146.0, 3.0), (147.0, 7.0), (148.0, 0.0), (149.0, 2.0), (150.0, 4.0), (151.0, 6.0), (152.0, 3.0), (153.0, 1.0), (154.0, 82.0), (155.0, 330.0), (156.0, 9.0), (157.0, 19.0), (158.0, 5.0), (159.0, 4.0), (160.0, 7.0), (161.0, 4.0), (162.0, 1.0), (163.0, 1.0), (164.0, 6.0), (165.0, 10.0), (166.0, 3.0), (167.0, 8.0), (168.0, 4.0), (169.0, 1.0), (170.0, 1.0), (171.0, 2.0), (172.0, 3.0), (173.0, 0.0), (174.0, 6.0), (175.0, 1.0), (176.0, 5.0), (177.0, 2.0), (178.0, 0.0), (179.0, 2.0), (180.0, 2.0), (181.0, 2.0), (182.0, 2.0), (183.0, 5.0), (184.0, 4.0), (185.0, 1.0), (186.0, 4.0), (187.0, 4.0), (188.0, 6.0), (189.0, 1.0), (190.0, 5.0), (191.0, 0.0), (192.0, 4.0), (193.0, 1.0), (194.0, 1.0), (195.0, 2.0), (196.0, 2.0), (197.0, 1.0), (198.0, 1.0), (199.0, 2.0), (200.0, 5.0), (201.0, 13.0), (202.0, 4.0), (203.0, 2.0), (204.0, 5.0), (205.0, 24.0), (206.0, 1.0), (207.0, 1.0), (208.0, 1.0), (209.0, 1.0), (210.0, 2.0), (211.0, 3.0), (212.0, 3.0), (213.0, 2.0), (214.0, 0.0), (215.0, 0.0), (216.0, 0.0), (217.0, 0.0), (218.0, 5.0), (219.0, 2.0), (220.0, 1.0), (221.0, 1.0), (222.0, 1.0), (223.0, 0.0), (224.0, 1.0)

UNITS: Tweets

"Daily_HashTag_with_Moving_Average_(7_days)" = GRAPH(TIME)

Points: (0.0, 0.0), (1.0, 0.0), (2.0, 0.0), (3.0, 0.0), (4.0, 0.0), (5.0, 0.0), (6.0, 0.0), (7.0, 0.0), (8.0, 0.0), (9.0, 0.0), (10.0, 0.0), (11.0, 0.0), (12.0, 0.0), (13.0, 0.0), (14.0, 0.0), (15.0, 0.0), (16.0, 0.0), (17.0, 0.0), (18.0, 0.0), (19.0, 0.0), (20.0, 0.0), (21.0, 0.0), (22.0, 0.0), (23.0, 18.7), (24.0, 19.4), (25.0, 22.1), (26.0, 25.6), (27.0, 29.9), (28.0, 36.1), (29.0, 37.4), (30.0, 19.4), (31.0, 19.9), (32.0, 17.7), (33.0, 14.6), (34.0, 10.4), (35.0, 6.4), (36.0, 5.1), (37.0, 4.7), (38.0, 3.7), (39.0, 3.1), (40.0, 2.9), (41.0, 2.7), (42.0, 0.6), (43.0, 0.6), (44.0, 0.7), (45.0, 0.6), (46.0, 1.1), (47.0, 1.3), (48.0, 1.6), (49.0, 1.7), (50.0, 1.7), (51.0, 1.3), (52.0, 1.4), (53.0, 0.9), (54.0, 1.0), (55.0, 0.9), (56.0, 1.0), (57.0, 1.0), (58.0, 1.0), (59.0, 1.3), (60.0, 2.9), (61.0, 3.7), (62.0, 4.1), (63.0, 4.0), (64.0, 4.0), (65.0, 4.6), (66.0, 4.3), (67.0, 3.3), (68.0, 2.3), (69.0, 1.7), (70.0, 2.0), (71.0, 2.3), (72.0, 1.7), (73.0, 2.6), (74.0, 2.4), (75.0, 2.3), (76.0, 3.3), (77.0, 5.9), (78.0, 7.4), (79.0, 11.3), (80.0, 13.1), (81.0, 18.0), (82.0, 20.3), (83.0, 22.9), (84.0, 23.7), (85.0, 26.0), (86.0, 25.9), (87.0, 29.6), (88.0, 33.4), (89.0, 40.4), (90.0, 68.7), (91.0, 109.0), (92.0, 141.9), (93.0, 183.7), (94.0, 206.0), (95.0, 245.4), (96.0, 256.0), (97.0, 240.9), (98.0, 205.9), (99.0, 185.1), (100.0, 158.0), (101.0, 150.0), (102.0, 114.9), (103.0, 108.6), (104.0, 109.4), (105.0, 108.4), (106.0, 101.3), (107.0, 94.0), (108.0, 82.4), (109.0, 74.7), (110.0, 67.1), (111.0, 64.3), (112.0, 62.9), (113.0, 62.0), (114.0, 56.1), (115.0, 51.4), (116.0, 56.0), (117.0, 54.9), (118.0, 43.4), (119.0, 43.4), (120.0, 49.9), (121.0, 53.7), (122.0, 58.1), (123.0, 55.3), (124.0, 54.9), (125.0, 56.0), (126.0, 52.1), (127.0, 40.1), (128.0, 34.6), (129.0, 28.0), (130.0, 21.6), (131.0, 19.7), (132.0, 18.6), (133.0, 17.9), (134.0, 17.9), (135.0, 15.0), (136.0, 14.6), (137.0, 14.7), (138.0, 14.1), (139.0, 11.9), (140.0, 11.7), (141.0, 13.6), (142.0, 17.3), (143.0, 16.1), (144.0, 15.6), (145.0, 16.3), (146.0, 15.7), (147.0, 14.6), (148.0, 10.0), (149.0, 5.9), (150.0, 5.6), (151.0, 5.9), (152.0, 3.6), (153.0, 3.3), (154.0, 14.0), (155.0, 61.1), (156.0, 62.1), (157.0, 64.3), (158.0, 64.1), (159.0, 64.3), (160.0, 65.1), (161.0, 54.0), (162.0, 7.0), (163.0, 5.9), (164.0, 4.0), (165.0, 4.7), (166.0, 4.6), (167.0, 4.7), (168.0, 4.7), (169.0, 4.7), (170.0, 4.7), (171.0, 4.1), (172.0, 3.1), (173.0, 2.7), (174.0, 2.4), (175.0, 2.0), (176.0, 2.6), (177.0, 2.7), (178.0, 2.4), (179.0, 2.3), (180.0, 2.6), (181.0, 2.0), (182.0, 2.1), (183.0, 2.1), (184.0, 2.4), (185.0, 2.6), (186.0, 2.9), (187.0, 3.1), (188.0, 3.7), (189.0, 3.6), (190.0, 3.6), (191.0, 3.0), (192.0, 3.4), (193.0, 3.0), (194.0, 2.6), (195.0, 2.0), (196.0, 2.1), (197.0, 1.6), (198.0, 1.7), (199.0, 1.4), (200.0, 2.0), (201.0, 3.7), (202.0, 4.0), (203.0, 4.0), (204.0, 4.6), (205.0, 7.9), (206.0, 7.7), (207.0, 7.1), (208.0, 5.4), (209.0, 5.0), (210.0, 5.0), (211.0, 4.7), (212.0, 1.7), (213.0, 1.9), (214.0, 1.7), (215.0, 1.6), (216.0, 1.4), (217.0, 1.1), (218.0, 1.4), (219.0, 1.3), (220.0, 1.1), (221.0, 1.3), (222.0, 1.4), (223.0, 1.4), (224.0, 1.6)

UNITS: Tweets/day

Disbeliever_Active_Percentage = Disbeliever_Active*100/Total

UNITS: fraction

Disbeliever_Adoption_Time = 1

UNITS: Days

Disbeliever_Dormant_Percentage = Disbeliever_Dormant*100/Total

UNITS: fraction

Disbeliever_Quit_Time = 1/0.11

UNITS: Days

Effect_of_Corrective_Info_on_Believer_Activation_Fraction =
GRAPH((Corrective_info_per_capita)/Standard_corrective_info_per_capita)

Points: (0.000, 0.8000), (0.250, 0.8019), (0.500, 0.8114), (0.750, 0.8564), (1.000, 1.0000), (1.250, 1.0806), (1.500, 1.1261), (1.750, 1.1526), (2.000, 1.1678), (2.250, 1.1754), (2.500, 1.1810), (2.750, 1.1848), (3.000, 1.1867), (3.250, 1.1894), (3.500, 1.1905), (3.750, 1.1924), (4.000, 1.1943), (4.250, 1.1962), (4.500, 1.1962), (4.750, 1.1981), (5.000, 1.2000) {GF EXTRAPOLATED}

UNITS: Dimensionless

Effect_of_Corrective_Info_on_Prob_of_False_Persuasion =
GRAPH(Corrective_info_per_capita/Standard_corrective_info_per_capita)

Points: (0.000, -0.000669285092428), (0.500, -0.00179862099621), (1.000, -0.00474258731776), (1.500, -0.0119202922022), (2.000, -0.026894142137), (2.500, -0.05), (3.000, -0.073105857863), (3.500, -0.0880797077978), (4.000, -0.0952574126822), (4.500, -0.0982013790038), (5.000, -0.0993307149076) {GF EXTRAPOLATED}

UNITS: Dimensionless

Effect_of_Misinformation_on_Disbeliever_Activation_Fraction =
GRAPH((Misinformation_per_capita/Standard_misinformation_per_capita))

Points: (0.000, 0.000), (0.250, 0.035), (0.500, 0.159), (0.750, 0.493), (1.000, 1.000), (1.250, 1.366), (1.500, 1.542), (1.750, 1.656), (2.000, 1.727), (2.250, 1.771), (2.500, 1.806), (2.750, 1.822), (3.000, 1.844), (3.250, 1.865), (3.500, 1.886), (3.750, 1.900), (4.000, 1.908), (4.250, 1.915), (4.500, 1.922), (4.750, 1.929), (5.000, 1.936) {GF EXTRAPOLATED}

UNITS: Dimensionless

Effect_of_Misinformation_on_Prob_of_False_Persuasion =
GRAPH(Misinformation_per_capita/Standard_misinformation_per_capita)

Points: (0.000, 0.000669285092428), (0.500, 0.00179862099621), (1.000, 0.00474258731776), (1.500, 0.0119202922022), (2.000, 0.026894142137), (2.500, 0.05),

(3.000, 0.073105857863), (3.500, 0.0880797077978), (4.000, 0.0952574126822), (4.500, 0.0982013790038), (5.000, 0.0993307149076) {GF EXTRAPOLATED}

UNITS: Dimensionless

Exposed_Percentage = Exposed*100/Total

UNITS: fraction

Exposure_Percentage = 100*(Total-Susceptible)/Total

UNITS: fraction

Google_Trends_Data = GRAPH(TIME)

Points: (0.0, 0.0), (1.0, 0.0), (2.0, 0.0), (3.0, 0.0), (4.0, 0.0), (5.0, 0.0), (6.0, 0.0), (7.0, 0.0), (8.0, 0.0), (9.0, 0.0), (10.0, 0.0), (11.0, 0.0), (12.0, 0.0), (13.0, 0.0), (14.0, 0.0), (15.0, 0.0), (16.0, 0.0), (17.0, 0.0), (18.0, 0.0), (19.0, 0.0), (20.0, 0.0), (21.0, 0.0), (22.0, 0.0), (23.0, 0.0), (24.0, 0.0), (25.0, 1.0), (26.0, 0.0), (27.0, 0.0), (28.0, 1.0), (29.0, 0.0), (30.0, 0.0), (31.0, 0.0), (32.0, 0.0), (33.0, 1.0), (34.0, 1.0), (35.0, 0.0), (36.0, 0.0), (37.0, 0.0), (38.0, 1.0), (39.0, 1.0), (40.0, 0.0), (41.0, 0.0), (42.0, 0.0), (43.0, 1.0), (44.0, 0.0), (45.0, 0.0), (46.0, 0.0), (47.0, 0.0), (48.0, 0.0), (49.0, 0.0), (50.0, 1.0), (51.0, 1.0), (52.0, 1.0), (53.0, 1.0), (54.0, 2.0), (55.0, 2.0), (56.0, 2.0), (57.0, 2.0), (58.0, 2.0), (59.0, 1.0), (60.0, 2.0), (61.0, 3.0), (62.0, 3.0), (63.0, 2.0), (64.0, 4.0), (65.0, 4.0), (66.0, 11.0), (67.0, 9.0), (68.0, 14.0), (69.0, 16.0), (70.0, 13.0), (71.0, 14.0), (72.0, 17.0), (73.0, 14.0), (74.0, 13.0), (75.0, 12.0), (76.0, 8.0), (77.0, 10.0), (78.0, 9.0), (79.0, 7.0), (80.0, 11.0), (81.0, 13.0), (82.0, 16.0), (83.0, 19.0), (84.0, 20.0), (85.0, 24.0), (86.0, 30.0), (87.0, 42.0), (88.0, 72.0), (89.0, 76.0), (90.0, 66.0), (91.0, 100.0), (92.0, 58.0), (93.0, 36.0), (94.0, 29.0), (95.0, 23.0), (96.0, 19.0), (97.0, 15.0), (98.0, 12.0), (99.0, 10.0), (100.0, 20.0), (101.0, 20.0), (102.0, 15.0), (103.0, 7.0), (104.0, 6.0), (105.0, 5.0), (106.0, 7.0), (107.0, 6.0), (108.0, 6.0), (109.0, 4.0), (110.0, 5.0), (111.0, 5.0), (112.0, 4.0), (113.0, 4.0), (114.0, 3.0), (115.0, 3.0), (116.0, 2.0), (117.0, 2.0), (118.0, 2.0), (119.0, 3.0), (120.0, 6.0), (121.0, 4.0), (122.0, 3.0), (123.0, 2.0), (124.0, 2.0), (125.0, 3.0), (126.0, 2.0), (127.0, 2.0), (128.0, 3.0), (129.0, 1.0), (130.0, 2.0), (131.0, 3.0), (132.0, 2.0), (133.0, 5.0), (134.0, 6.0), (135.0, 2.0), (136.0, 1.0), (137.0, 2.0), (138.0, 2.0), (139.0, 1.0), (140.0, 2.0), (141.0, 2.0), (142.0, 1.0), (143.0, 1.0), (144.0, 1.0), (145.0, 1.0), (146.0, 1.0), (147.0, 0.0), (148.0, 1.0), (149.0, 1.0), (150.0, 0.0), (151.0, 1.0), (152.0, 1.0), (153.0, 1.0), (154.0, 1.0), (155.0, 1.0), (156.0, 1.0), (157.0, 1.0), (158.0, 1.0), (159.0, 1.0), (160.0, 0.0), (161.0, 0.0), (162.0, 1.0), (163.0, 1.0), (164.0, 0.0), (165.0, 0.0), (166.0, 1.0), (167.0, 0.0), (168.0, 0.0), (169.0, 1.0), (170.0, 1.0), (171.0, 1.0), (172.0, 1.0), (173.0, 0.0), (174.0, 1.0), (175.0, 1.0), (176.0, 1.0), (177.0, 0.0), (178.0, 0.0), (179.0, 1.0), (180.0, 0.0), (181.0, 0.0), (182.0, 2.0), (183.0, 3.0), (184.0, 1.0), (185.0, 2.0), (186.0, 0.0), (187.0, 1.0), (188.0, 1.0), (189.0, 1.0), (190.0, 1.0), (191.0, 0.0), (192.0, 2.0), (193.0, 1.0), (194.0, 0.0), (195.0, 0.0), (196.0, 0.0), (197.0, 1.0), (198.0, 1.0), (199.0, 1.0), (200.0, 1.0), (201.0, 1.0), (202.0, 2.0), (203.0, 0.0), (204.0, 1.0), (205.0, 1.0),

(206.0, 0.0), (207.0, 0.0), (208.0, 0.0), (209.0, 0.0), (210.0, 1.0), (211.0, 0.0), (212.0, 1.0),
(213.0, 1.0), (214.0, 0.0), (215.0, 0.0), (216.0, 0.0), (217.0, 0.0), (218.0, 0.0), (219.0, 1.0),
(220.0, 0.0), (221.0, 0.0), (222.0, 0.0), (223.0, 0.0), (224.0, 0.0)

UNITS: Dimensionless

Informed_Incidence_Percentage = Total_Quit_from_Disbeliever*100/Total

UNITS: fraction

"IsSuperspreader_(bool)" = 0

UNITS: Dimensionless

Labeled_Cumulative_Monthy_Tweets = GRAPH(TIME)

Points: (15.0, 2), (45.0, 14), (75.0, 199), (105.0, 1141), (135.0, 1348), (165.0, 1430),
(195.0, 1501), (225.0, 1536), (255.0, 1595), (285.0, 1653), (315.0, 1711), (345.0, 1752),
(375.0, 1799), (405.0, 1819), (435.0, 1842), (465.0, 1864), (495.0, 1882), (525.0, 1902),
(555.0, 1918), (585.0, 1926), (615.0, 1933), (645.0, 1946), (675.0, 1957)

UNITS: Tweets/day

Labeled_Monthy_Tweet_Counts = GRAPH(TIME)

Points: (15.0, 2), (45.0, 12), (75.0, 186), (105.0, 942), (135.0, 208), (165.0, 82), (195.0,
71), (225.0, 35), (255.0, 59), (285.0, 57), (315.0, 58), (345.0, 41), (375.0, 47), (405.0, 20),
(435.0, 23), (465.0, 23), (495.0, 17), (525.0, 21), (555.0, 16), (585.0, 8), (615.0, 7),
(645.0, 13), (675.0, 11)

UNITS: Tweets/day

MAX_COVID_Case_4807 = 4807

UNITS: Case per day

Max_Tweet_Count = 942

UNITS: Tweets/day

Misinformation_Depreciation_Time = 2

UNITS: Days

Misinformation_per_capita = Misinformation/Total

UNITS: information/people

Neutral_Adoption_Time = 1

UNITS: Days

Neutral_Dormant_Percentage = Neutral*100/Total

UNITS: fraction

Neutral_Engagement_Fraction = 0

UNITS: fraction

Neutral_Fract = 0.1

UNITS: Dimensionless

Neutral_Incidence_Percentage = Total_Quit_from_Neutral*100/Total

UNITS: fraction

Neutral_Misinformation_Generation_Per_people = 1

UNITS: Information/(Day*People)

Neutral_Quit_Time = 1/0.11

UNITS: Days

Normal_Believer_Activation_Fraction = 0.68

UNITS: Per Day

Normal_Disbeliever_Activation_Fraction = 0.2

UNITS: Per Day

Normal_Prob_of_False_Persuasion = 0.22

UNITS: Dimensionless

"Normalized_Daily_HashTag_with_Moving_Average_(7_days)" =
"Daily_HashTag_with_Moving_Average_(7_days)"/256

UNITS: Tweets/day

Normalized_Google_Trends_Data = Google_Trends_Data/100

UNITS: Dimensionless

Normalized_Labeled_Monthly_Tweet_Data =
Labeled_Monthy_Tweet_Counts/Max_Tweet_Count

UNITS: Dimensionless

Normalized_UK_COVID_Cases = UK_COVID_Cases/MAX_COVID_Case_4807

UNITS: Dimensionless

Prob_of_False_Persuasion_Effect_Multiplier = 1

UNITS: Dimensionless

Proportion_of_Infected_People = Total_Active/Total

UNITS: unitless

S_Initial = 10000

UNITS: People

Standard_corrective_info_per_capita = 0.02

UNITS: information/people

Standard_misinformation_per_capita = 0.02

UNITS: information/people

"Super-spreader_misinformation_generation" = 800

UNITS: Information/Day

"Super-spreader_popularity_duration" = 3

UNITS: days

"Super-spreader_start_time" = 100

UNITS: days

Superspreadер_Contact_fraction = 0.02

UNITS: per day

Susceptible_Percentage = 100*Susceptible/Total

UNITS: fraction

Total = S_Initial+Active_Initial

UNITS: People

Total_Active =

Disbeliever_Active+Believer_Active+Neutral*Neutral_Engagement_Fraction

UNITS: People

Total_Active_Peak = MAXPEAK(Total_Active)

UNITS: People

Total_Active_Percentage = 100*Total_Active/Total

UNITS: fraction

Total_Believer = Believer_Active + Believer_Dormant {SUMMING CONVERTER}

UNITS: People

Total_Believer_Peak = MAXPEAK(Total_Believer)

UNITS: People

Total_Believer_Peak_Percentage = 100*Total_Believer_Peak/Total

UNITS: fraction

Total_Believer_Percentage = 100*Total_Believer/Total

UNITS: fraction

Total_Dormant_&_Active = Believer_Active + Believer_Dormant + Disbeliever_Active + Disbeliever_Dormant {SUMMING CONVERTER}

UNITS: People

UK_COVID_Cases = GRAPH(TIME)

Points: (0.0, 0), (1.0, 0), (2.0, 0), (3.0, 0), (4.0, 0), (5.0, 0), (6.0, 0), (7.0, 0), (8.0, 0), (9.0, 0), (10.0, 0), (11.0, 0), (12.0, 0), (13.0, 0), (14.0, 0), (15.0, 0), (16.0, 0), (17.0, 0), (18.0, 0), (19.0, 0), (20.0, 0), (21.0, 0), (22.0, 0), (23.0, 0), (24.0, 0), (25.0, 0.0126470618269), (26.0, 0.114971909656), (27.0, 0.141090418371), (28.0, 0.156599046352), (29.0, 0.508823759911), (30.0, 2.29128878003), (31.0, 2.83061760866), (32.0, 2.95197177222), (33.0, 2.88832311389), (34.0, 2.9069161251), (35.0, 3.17974322034), (36.0, 3.0002338247), (37.0, 1.54960378475), (38.0, 1.33504410612), (39.0,

1.4316141213), (40.0, 1.53239325606), (41.0, 1.51495002153), (42.0, 1.24640134525),
(43.0, 1.05366200444), (44.0, 0.931469092928), (45.0, 0.580024252136), (46.0,
0.329049627621), (47.0, 0.229163362136), (48.0, 0.34483328792), (49.0,
0.533145420002), (50.0, 1.03117659655), (51.0, 1.40701870017), (52.0,
1.75009110977), (53.0, 2.18841290187), (54.0, 3.08865791174), (55.0, 4.25740647528),
(56.0, 5.30588796367), (57.0, 7.35537689578), (58.0, 11.9029476692), (59.0,
18.6568757825), (60.0, 25.9602910237), (61.0, 33.0775358698), (62.0, 41.5801972704),
(63.0, 50.2065871257), (64.0, 58.2758431213), (65.0, 72.0303915829), (66.0,
98.5174603281), (67.0, 141.332139689), (68.0, 196.315224452), (69.0, 251.540586766),
(70.0, 300.63067388), (71.0, 352.323306985), (72.0, 413.426704469), (73.0,
484.0616617), (74.0, 539.554418902), (75.0, 531.509239353), (76.0, 477.099981813),
(77.0, 417.700899359), (78.0, 404.142163672), (79.0, 516.333171818), (80.0,
724.832212721), (81.0, 987.686413292), (82.0, 1338.98784339), (83.0, 1756.17521353),
(84.0, 2173.5182275), (85.0, 2557.13816104), (86.0, 2888.4669305), (87.0,
3189.89102669), (88.0, 3487.20205426), (89.0, 3765.51218634), (90.0, 4008.2256326),
(91.0, 4202.49584163), (92.0, 4351.16165132), (93.0, 4487.77639863), (94.0,
4615.8209109), (95.0, 4698.43130548), (96.0, 4718.66403196), (97.0, 4687.72106213),
(98.0, 4638.77354868), (99.0, 4586.73023894), (100.0, 4496.17258637), (101.0,
4375.53067127), (102.0, 4303.39672977), (103.0, 4316.55002852), (104.0,
4394.8114463), (105.0, 4505.89096741), (106.0, 4602.0576982), (107.0,
4662.68598786), (108.0, 4714.74438838), (109.0, 4764.8388476), (110.0,
4789.41752234), (111.0, 4771.85497794), (112.0, 4708.83443384), (113.0,
4639.79209214), (114.0, 4614.89389878), (115.0, 4618.05280321), (116.0,
4611.47420285), (117.0, 4591.36178241), (118.0, 4559.93033757), (119.0,
4511.37167416), (120.0, 4433.43571418), (121.0, 4311.15369644), (122.0,
4148.78980783), (123.0, 3964.16254376), (124.0, 3767.95622154), (125.0,
3566.76341221), (126.0, 3382.78150534), (127.0, 3257.54206152), (128.0,
3205.37342662), (129.0, 3182.49637053), (130.0, 3128.62199276), (131.0,
3028.11684757), (132.0, 2923.7484034), (133.0, 2858.0419944), (134.0,
2802.49668909), (135.0, 2715.35558496), (136.0, 2619.59539105), (137.0,
2546.66551904), (138.0, 2494.31035254), (139.0, 2447.28268674), (140.0,
2384.8718092), (141.0, 2307.36790071), (142.0, 2208.91622678), (143.0,
2082.28448972), (144.0, 1948.83229631), (145.0, 1827.94864905), (146.0,
1729.39164584), (147.0, 1657.27469553), (148.0, 1605.24238081), (149.0,
1563.29938058), (150.0, 1513.00786104), (151.0, 1445.63521826), (152.0,
1376.21978989), (153.0, 1311.55790599), (154.0, 1255.6232037), (155.0,
1205.77540124), (156.0, 1156.66314051), (157.0, 1110.10749858), (158.0,
1071.19156031), (159.0, 1041.36484315), (160.0, 1019.31466361), (161.0,
1014.12735395), (162.0, 1022.49155975), (163.0, 1028.52370516), (164.0,
1022.82598938), (165.0, 1009.5907627), (166.0, 995.557696022), (167.0,
988.519663443), (168.0, 974.406651735), (169.0, 949.369703193), (170.0,
923.738982943), (171.0, 893.594240643), (172.0, 830.472626771), (173.0,
735.083779401), (174.0, 679.052933822), (175.0, 645.796680406), (176.0,

628.805080824), (177.0, 605.320465552), (178.0, 580.959622497), (179.0, 581.110921071), (180.0, 624.300951789), (181.0, 621.698335207), (182.0, 600.156269112), (183.0, 560.74647607), (184.0, 517.709643915), (185.0, 508.668446176), (186.0, 509.713837121), (187.0, 518.763631403), (188.0, 530.273077491), (189.0, 535.27319027), (190.0, 552.616099039), (191.0, 588.279549801), (192.0, 599.610892913), (193.0, 611.051427341), (194.0, 618.337901687), (195.0, 618.824733805), (196.0, 621.513488468), (197.0, 628.681916131), (198.0, 640.066152173), (199.0, 652.492381918), (200.0, 663.292374615), (201.0, 671.011531051), (202.0, 686.154914976), (203.0, 708.398705272), (204.0, 715.873980343), (205.0, 729.573127498), (206.0, 736.704045636), (207.0, 749.60392053), (208.0, 780.615843739), (209.0, 797.384567484), (210.0, 790.147506935), (211.0, 793.023710195), (212.0, 796.274159739), (213.0, 822.223815495), (214.0, 849.535903166), (215.0, 856.506997559), (216.0, 879.602570674), (217.0, 917.014762652), (218.0, 938.381936054), (219.0, 950.459439123), (220.0, 1010.74158138), (221.0, 1054.83182867), (222.0, 1076.20037107), (223.0, 1084.04207525), (224.0, 1113.72761919), (225.0, 1114.428571)

UNITS: Case per day

{ The model has 115 (115) variables (array expansion in parens).

In root model and 0 additional modules with 3 sectors.

Stocks: 15 (15) Flows: 24 (24) Converters: 76 (76)

Constants: 35 (35) Equations: 65 (65) Graphicals: 11 (11)

There are also 4 expanded macro variables.

}

APPENDIX B: ADDITIONAL EXTREME CONDITION TEST RESULTS

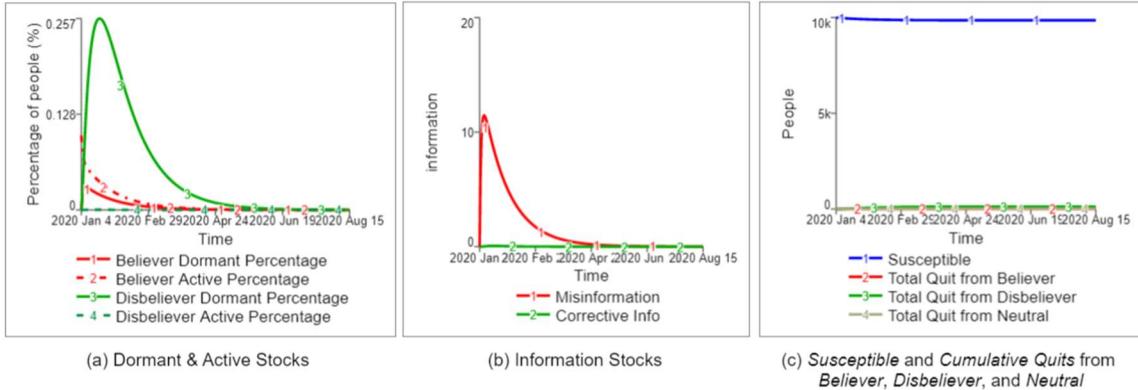


Figure B.1. Extreme condition (Normal Probability of False Persuasion = 0.001)

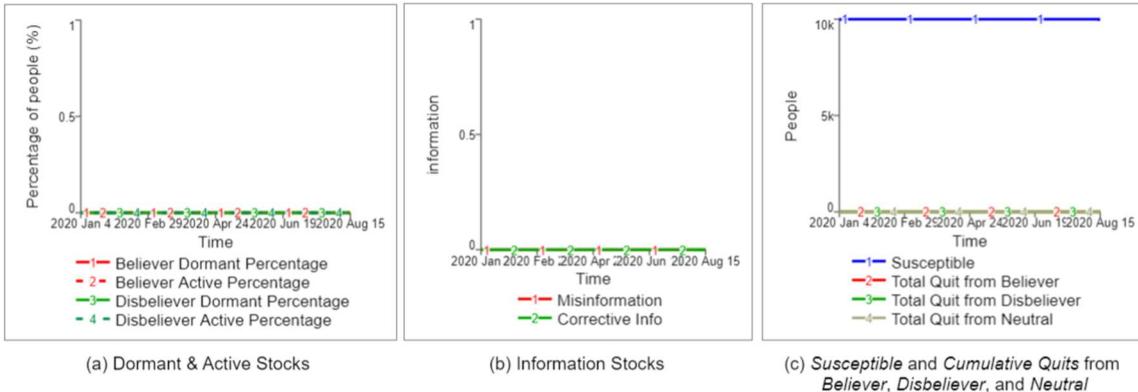


Figure B.2. Extreme condition (Active Initial = 0)

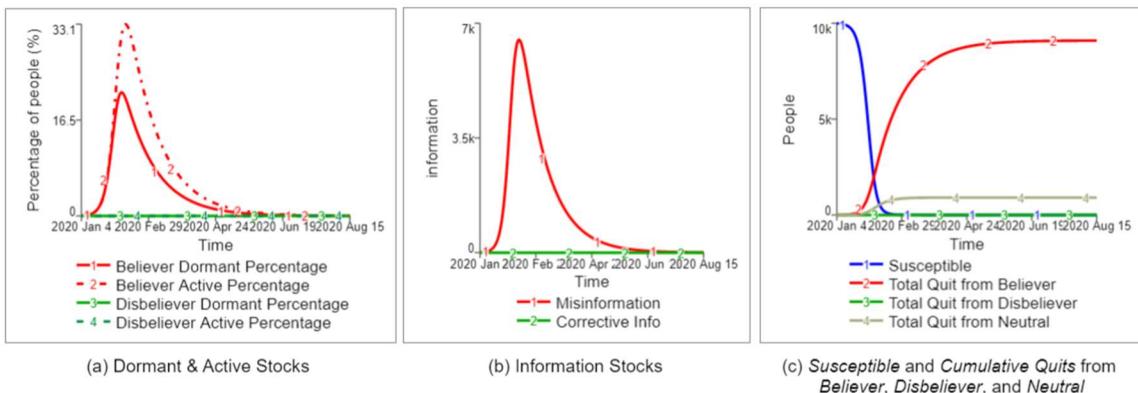


Figure B.3. Extreme condition (Normal Probability of False Persuasion = 0.999)

APPENDIX C: SENSITIVITY RUNS

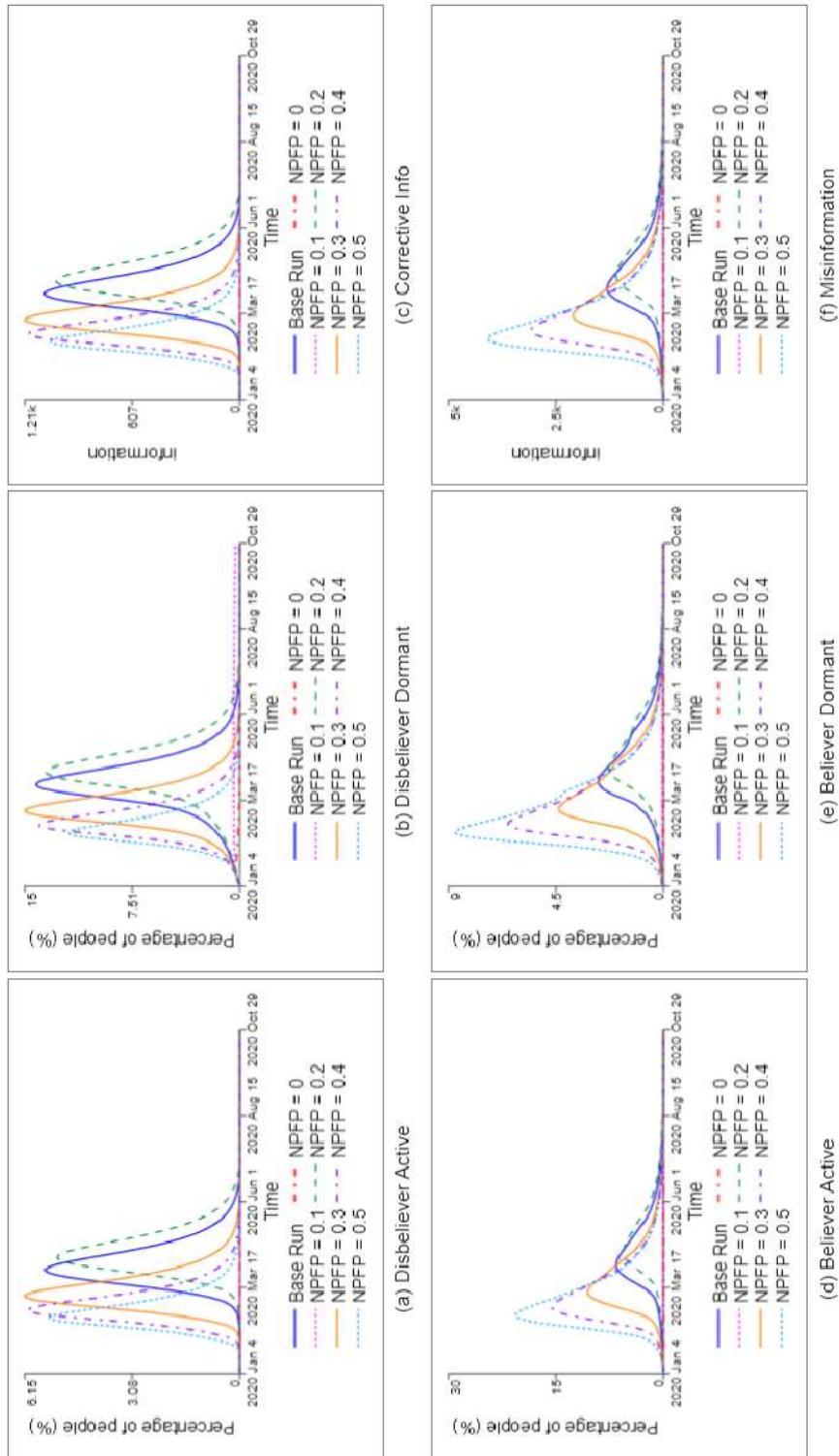


Figure C.1. Sensitivity to *Normal Probability of False Persuasion* (Base = 0.22)
(Panel 1)

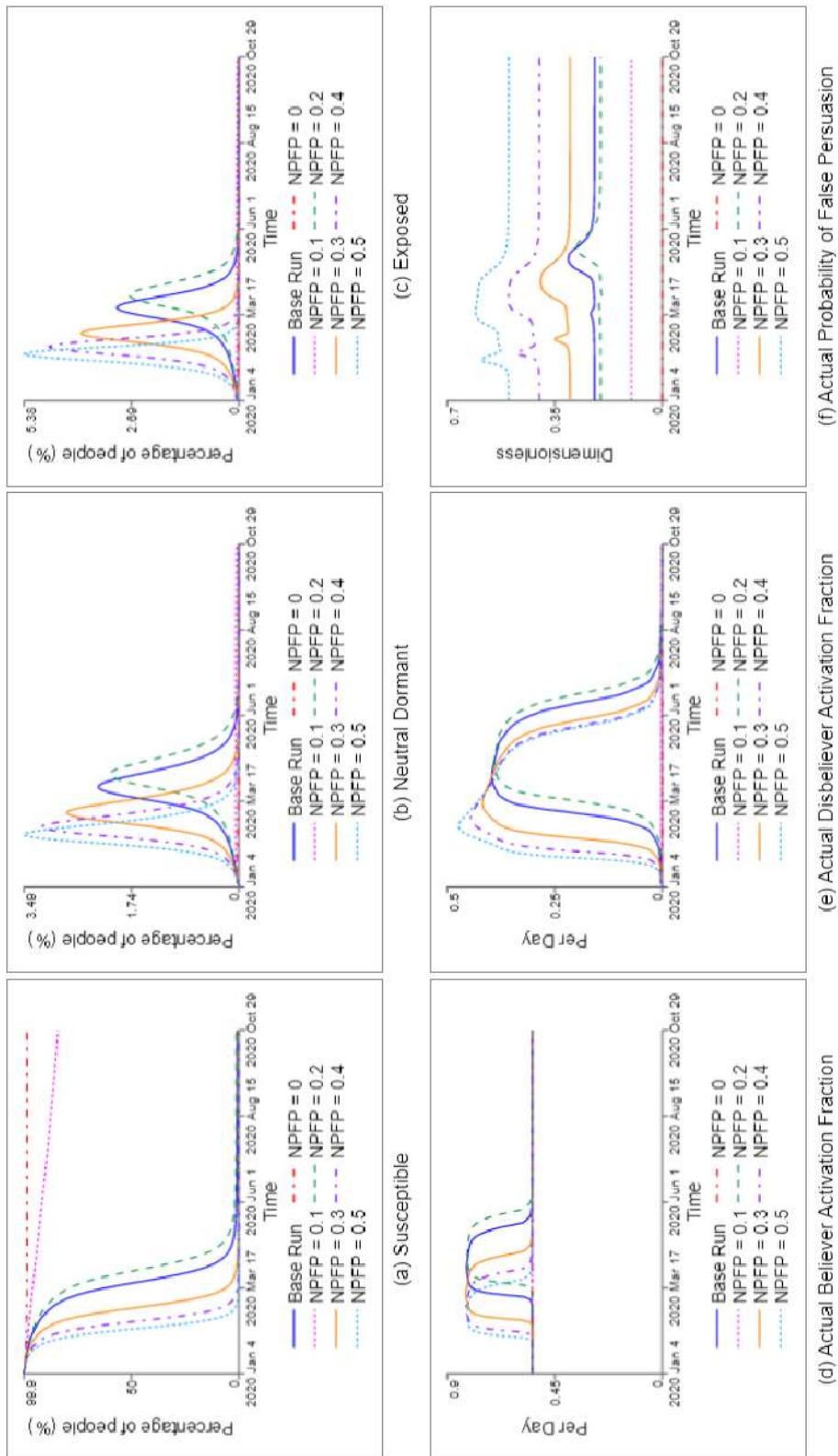


Figure C.2. Sensitivity to *Normal Probability of False Persuasion* (Base = 0.22)
(Panel 2)

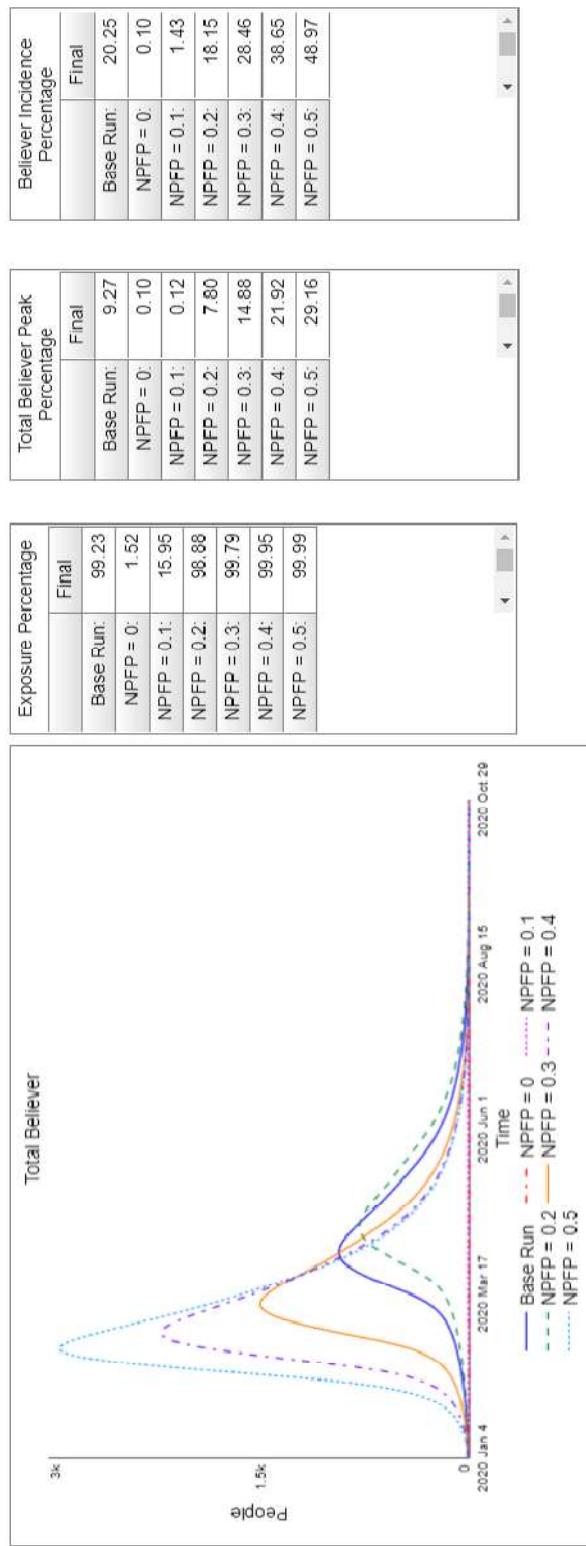


Figure C.3. Sensitivity to *Normal Probability of False Persuasion* (Base = 0.22)
(Panel 3)

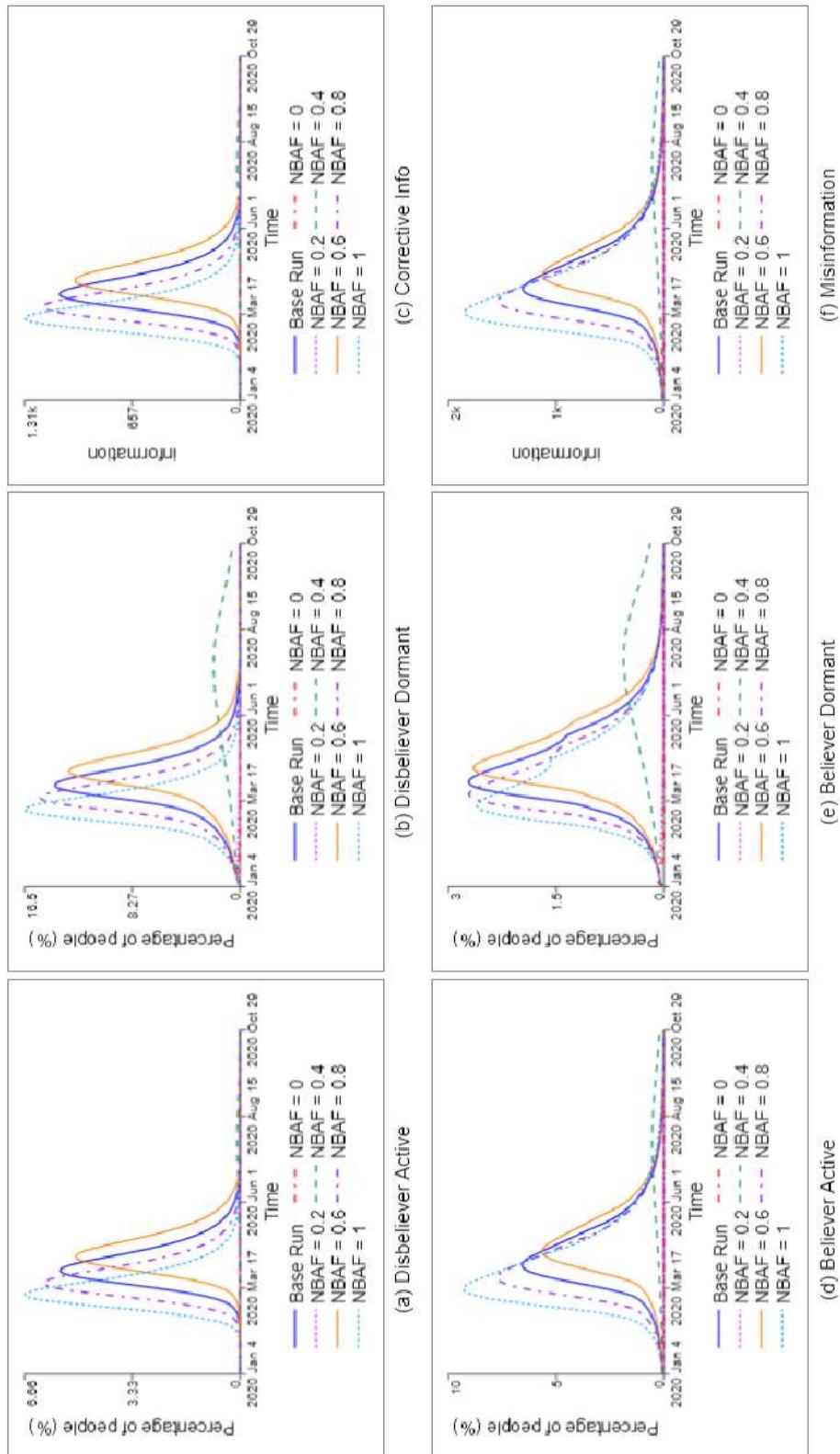


Figure C.4. Sensitivity to *Normal Believer Activation Fraction* (Base = 0.68)
(Panel 1)

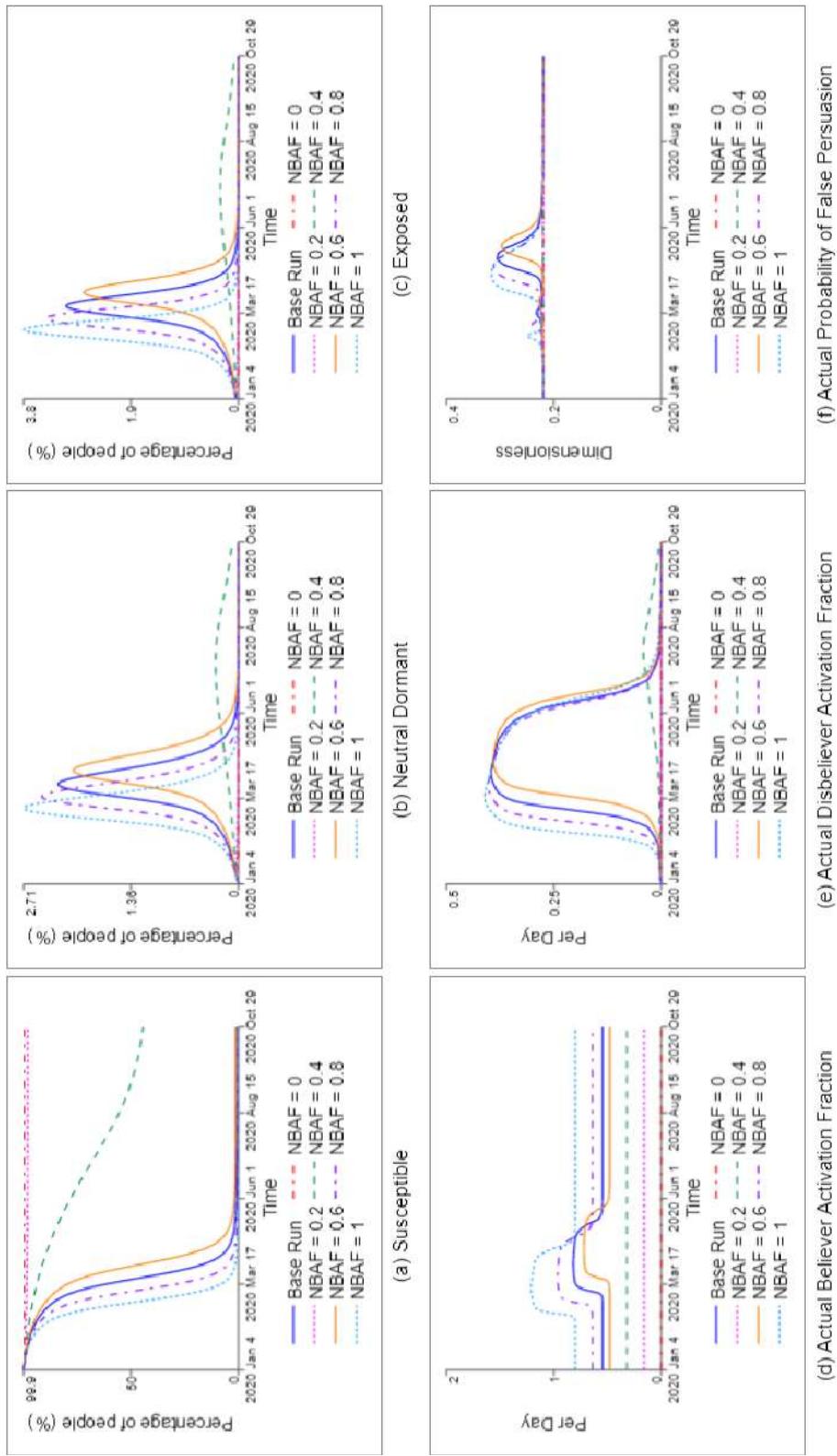


Figure C.5. Sensitivity to *Normal Believer Activation Fraction* (Base = 0.68)
(Panel 2)

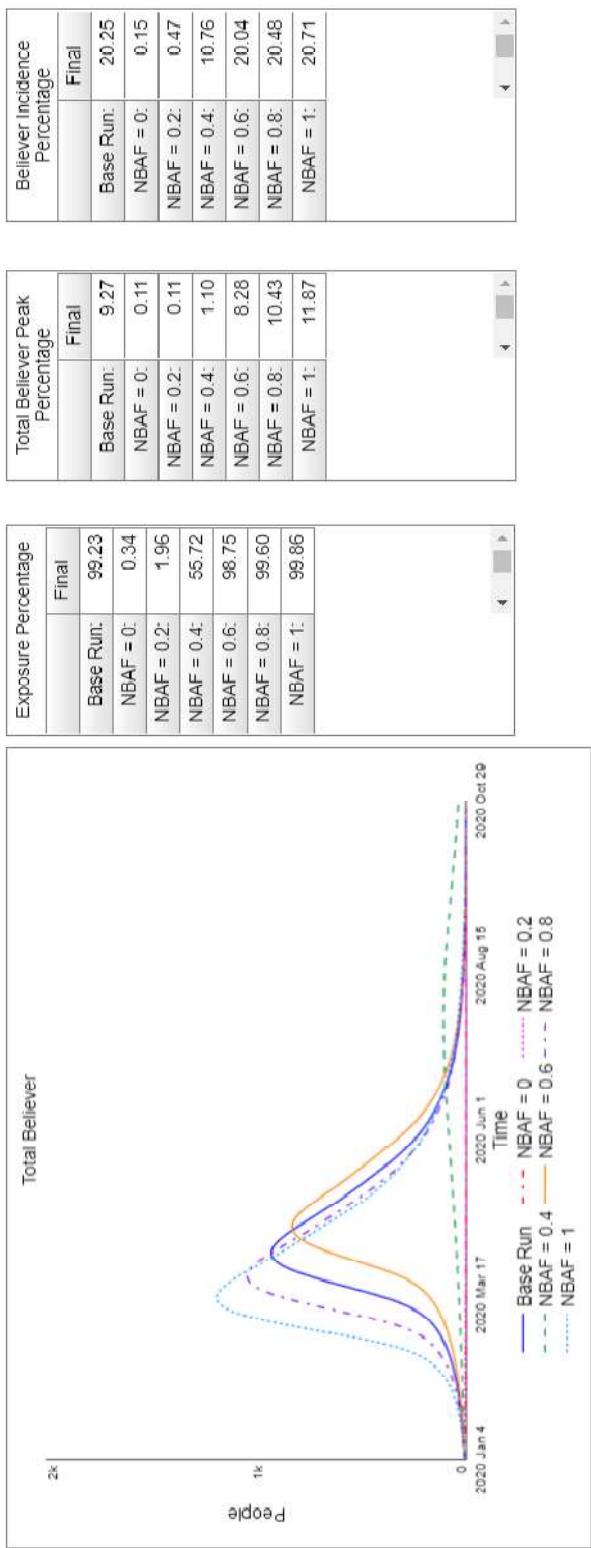


Figure C.6. Sensitivity to *Normal Believer Activation Fraction* (Base = 0.68)
(Panel 3)

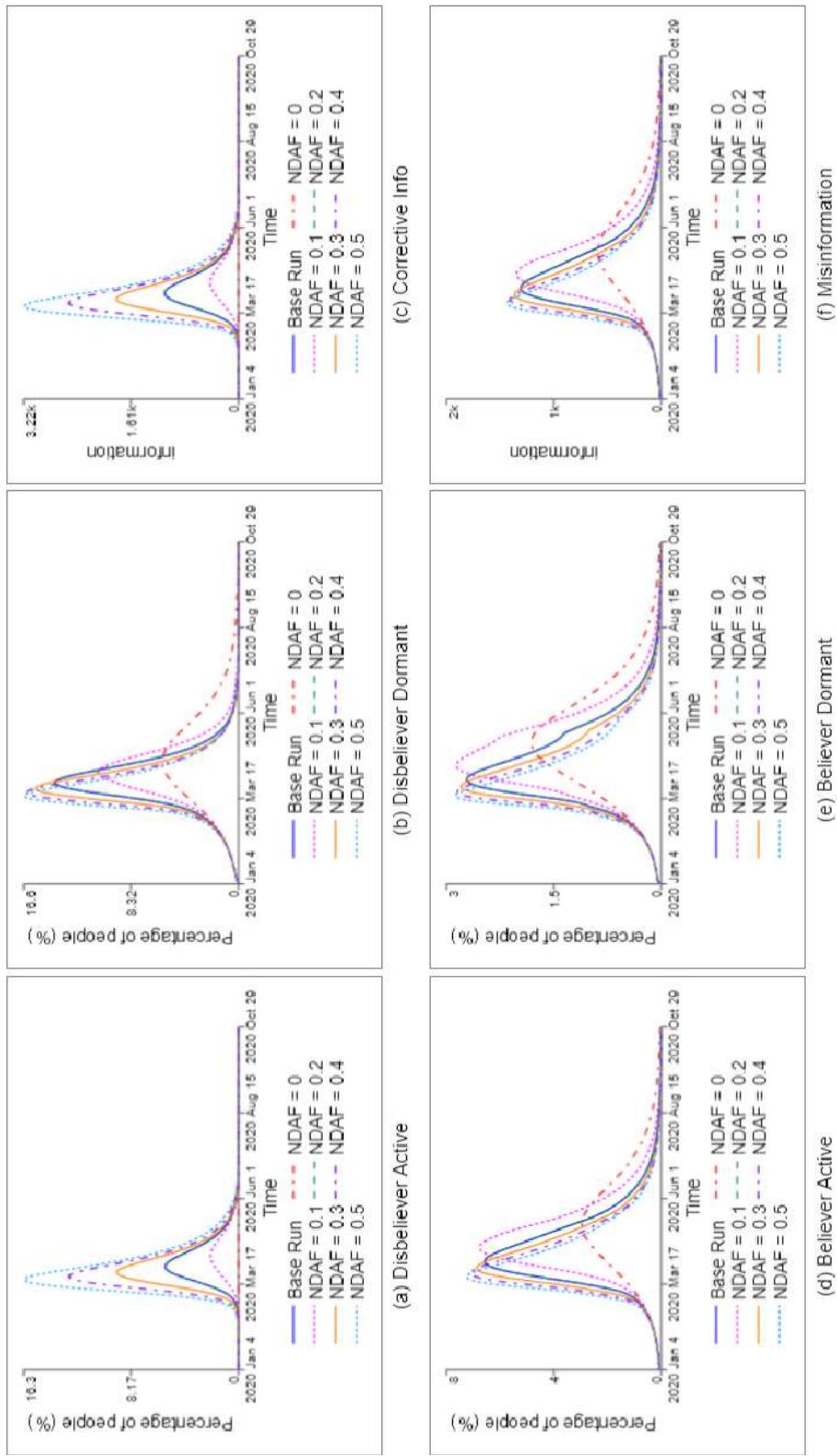


Figure C.7. Sensitivity to *Normal Disbeliever Activation Fraction* (Base = 0.2)
(Panel 1)

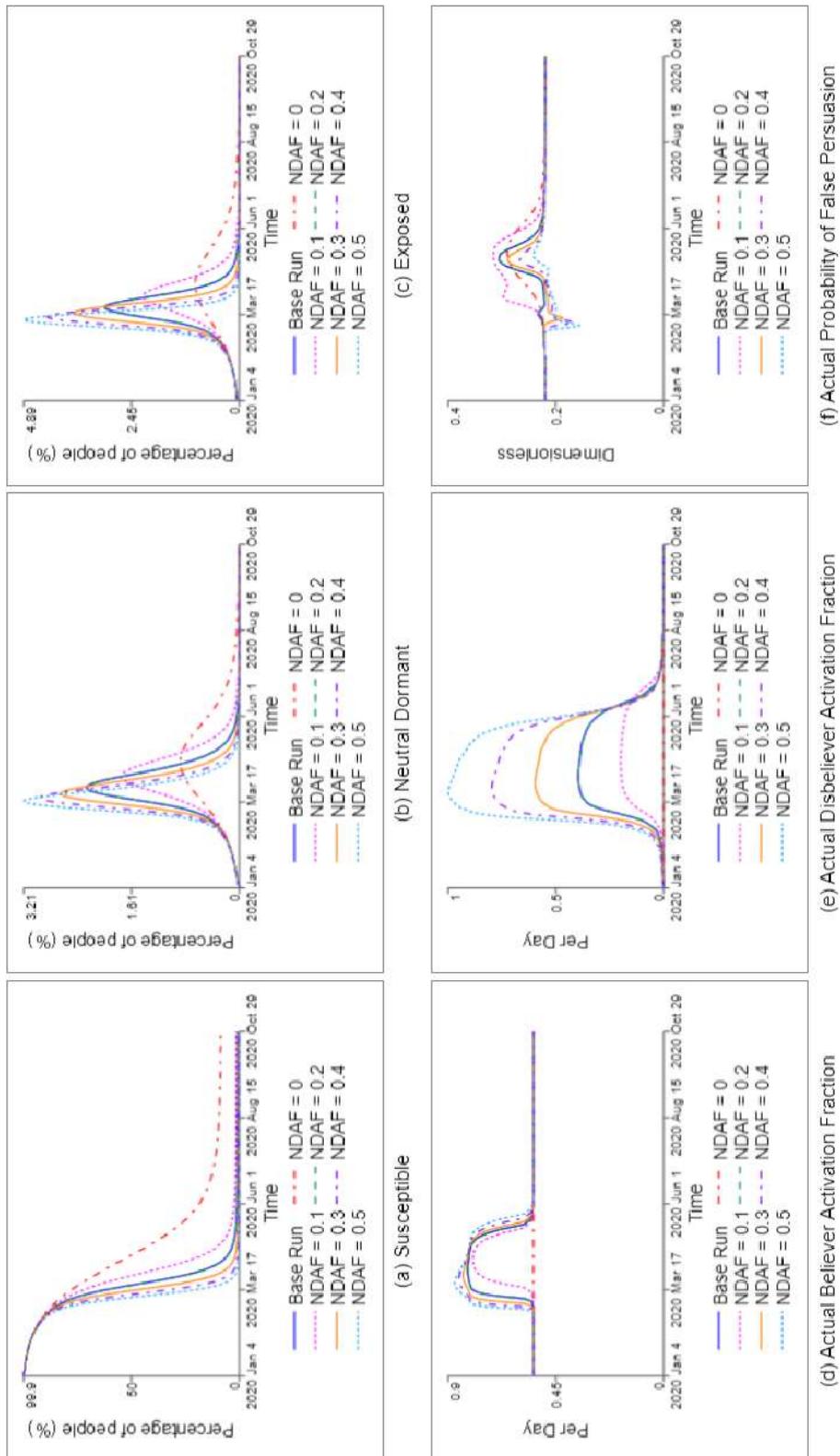


Figure C.8. Sensitivity to *Normal Disbeliever Activation Fraction* (Base = 0.2)
(Panel 2)

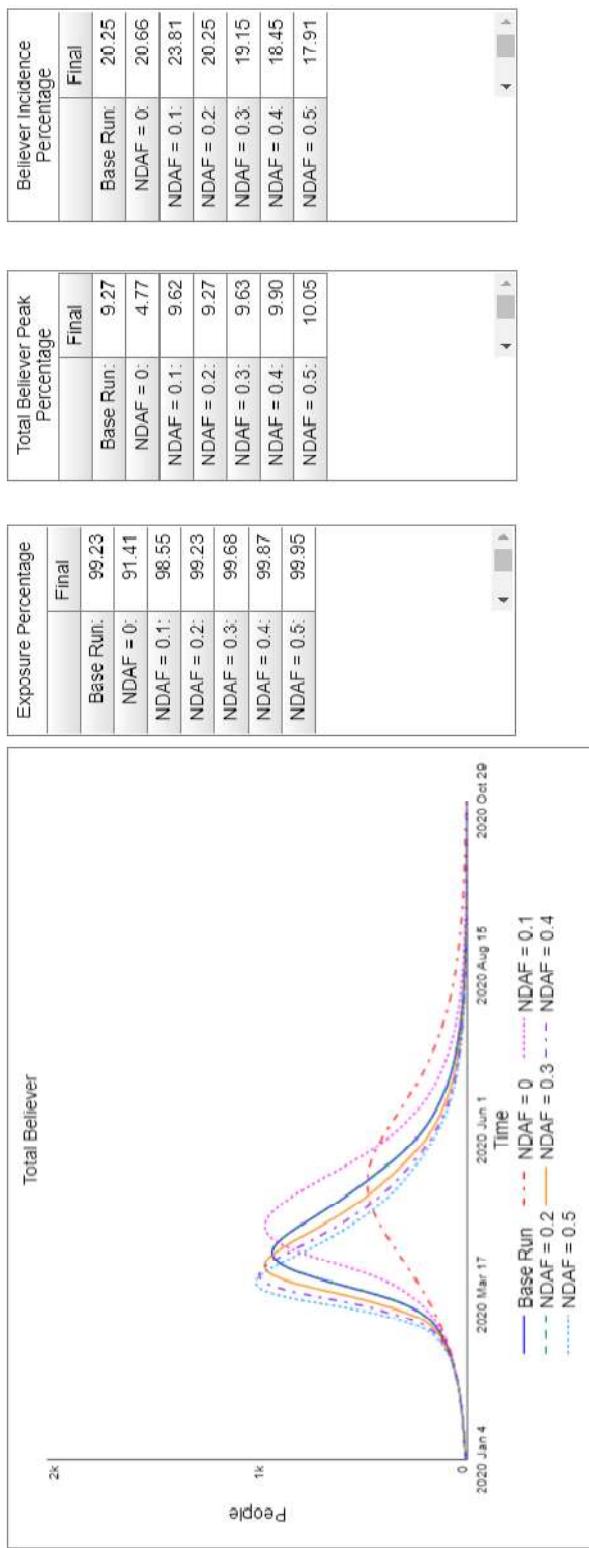


Figure C.9. Sensitivity to *Normal Disbeliever Activation Fraction* (Base = 0.2)
(Panel 3)

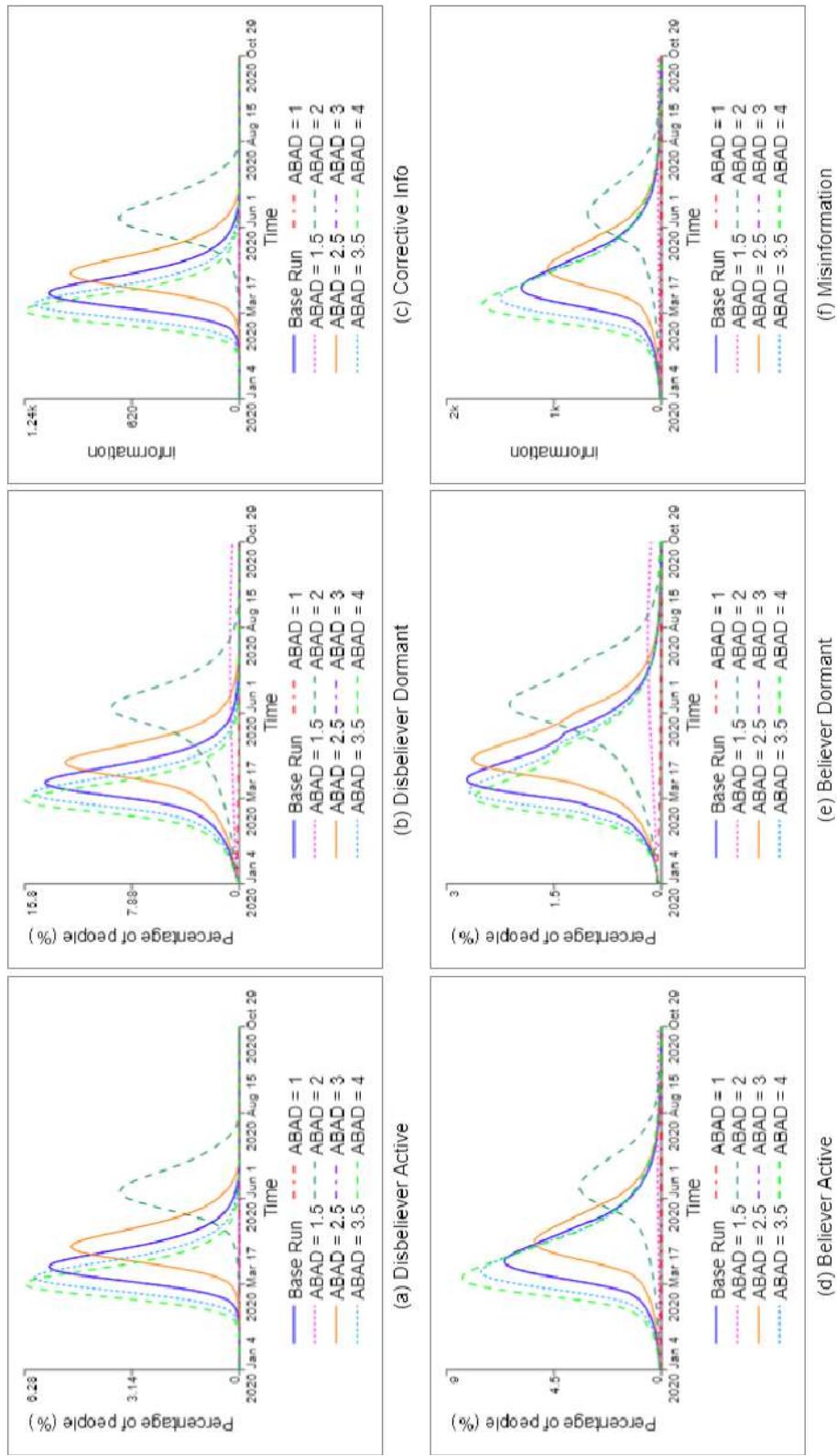


Figure C.10. Sensitivity to *Average Believer Active Duration* (Base = 3)
(Panel 1)

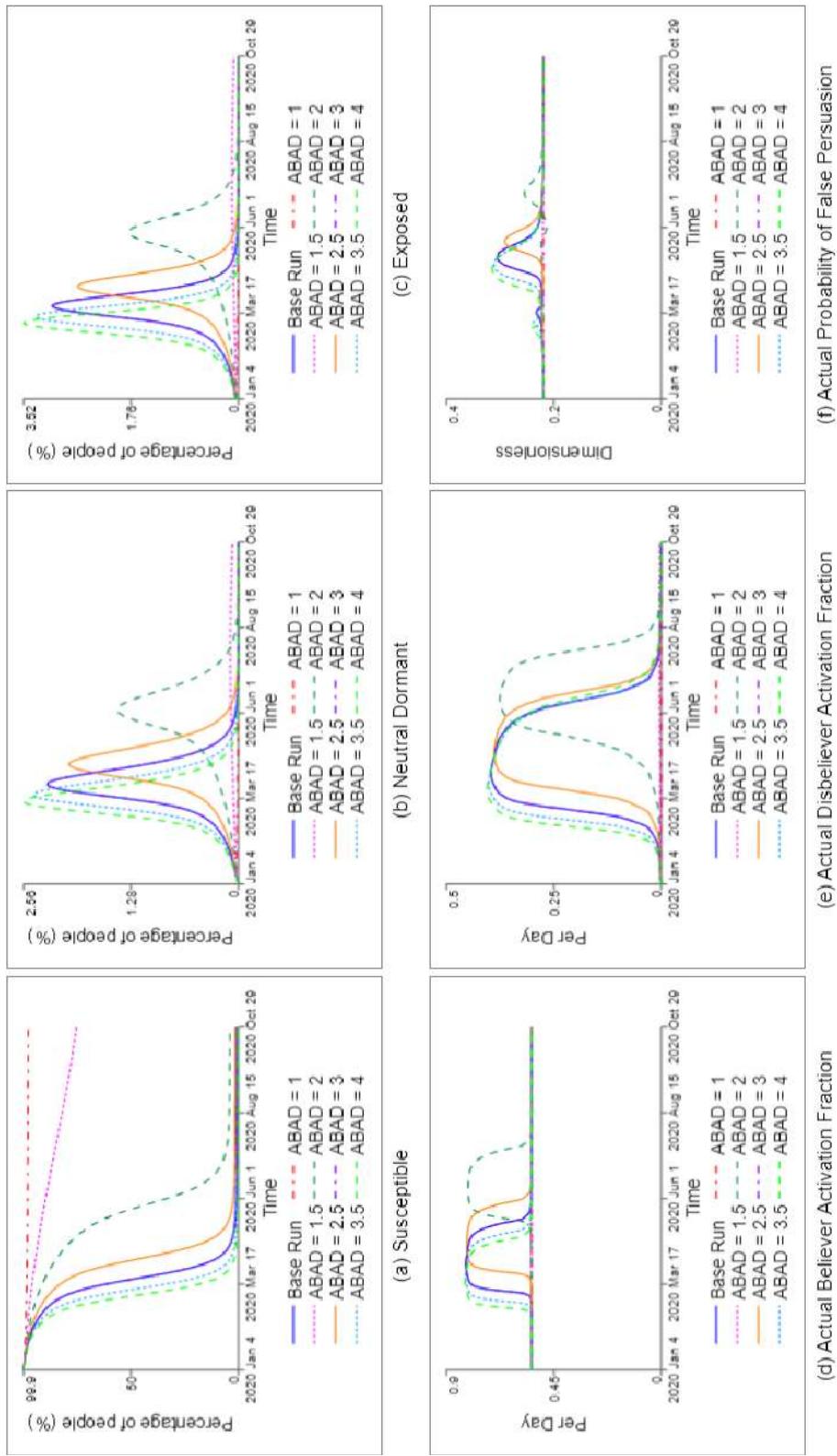


Figure C.11. Sensitivity to *Average Believer Active Duration* (Base = 3)
(Panel 2)

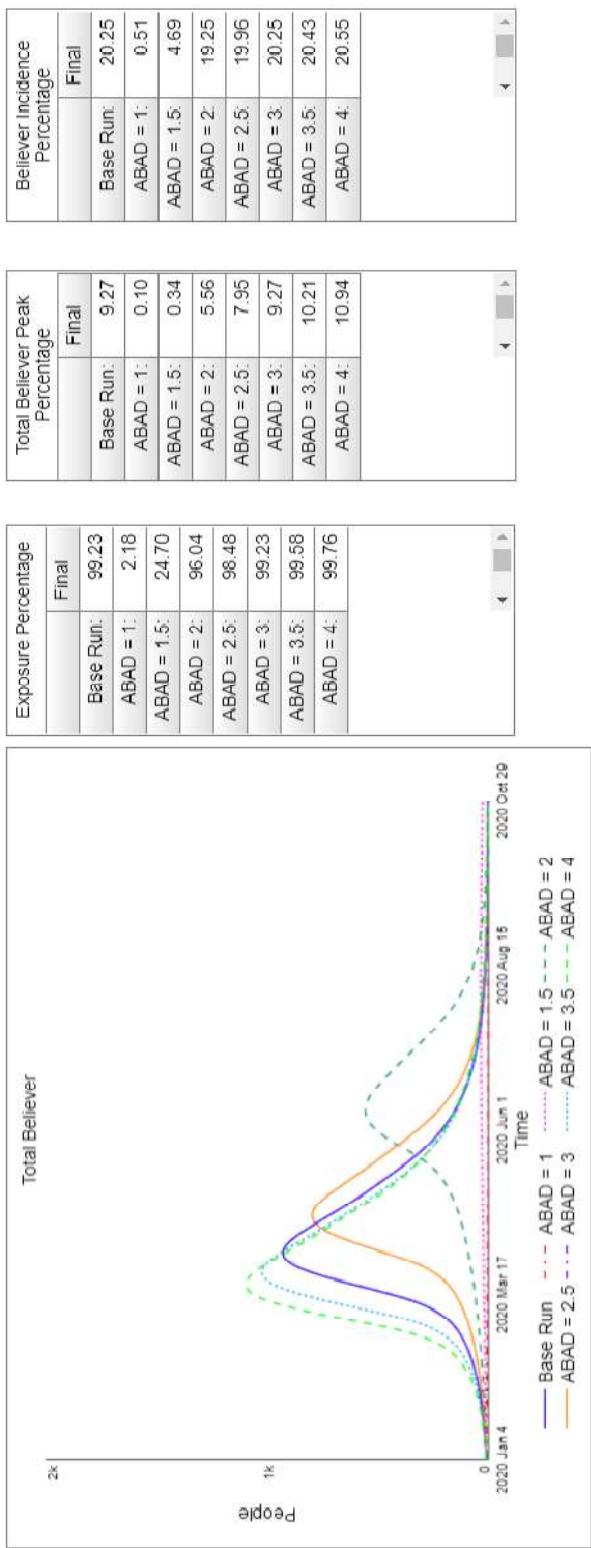


Figure C.12. Sensitivity to *Average Believer Active Duration* (Base = 3)
(Panel 3)

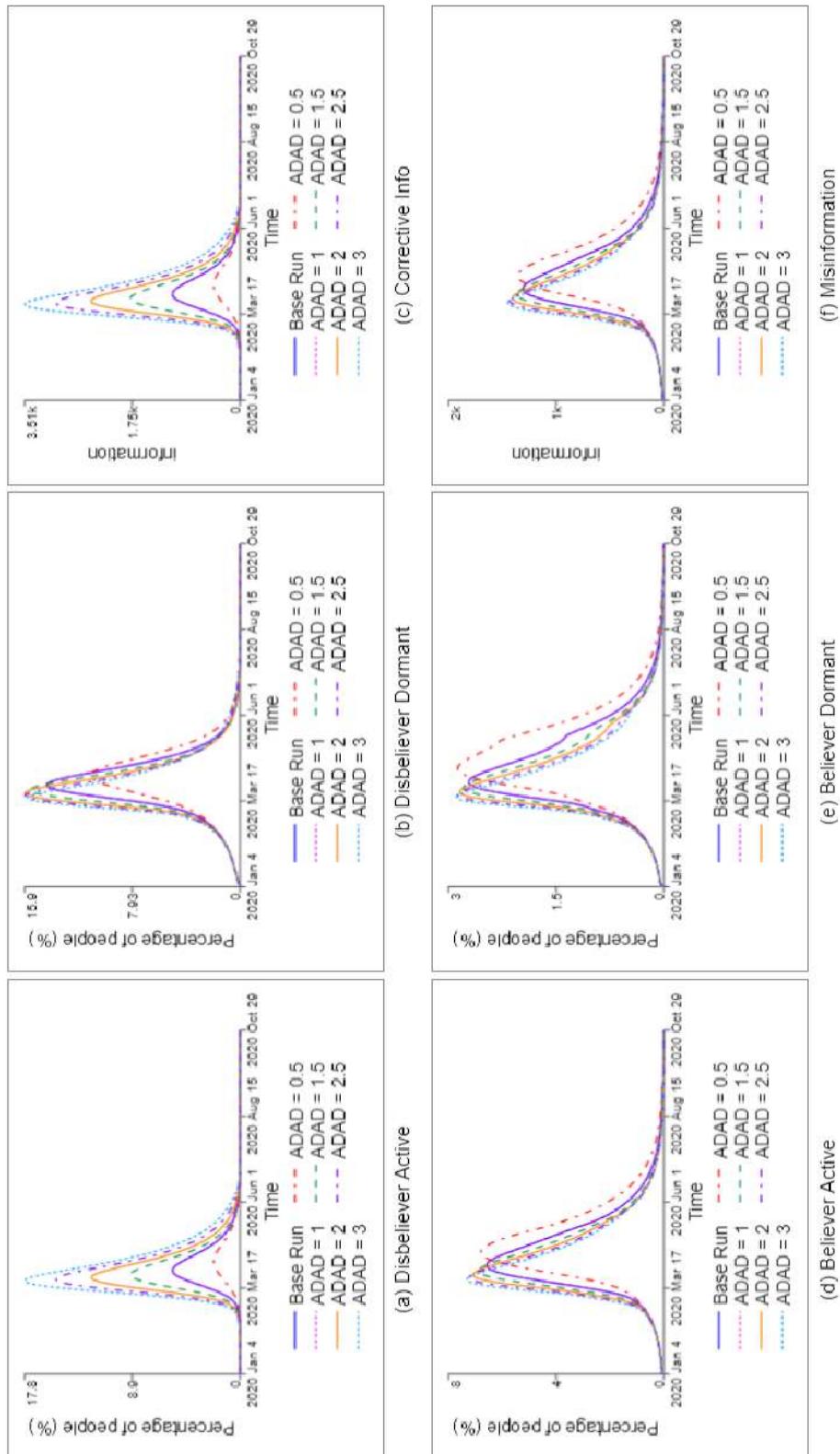


Figure C.13. Sensitivity to *Average Disbeliever Active Duration* (Base = 1)
(Panel 1)

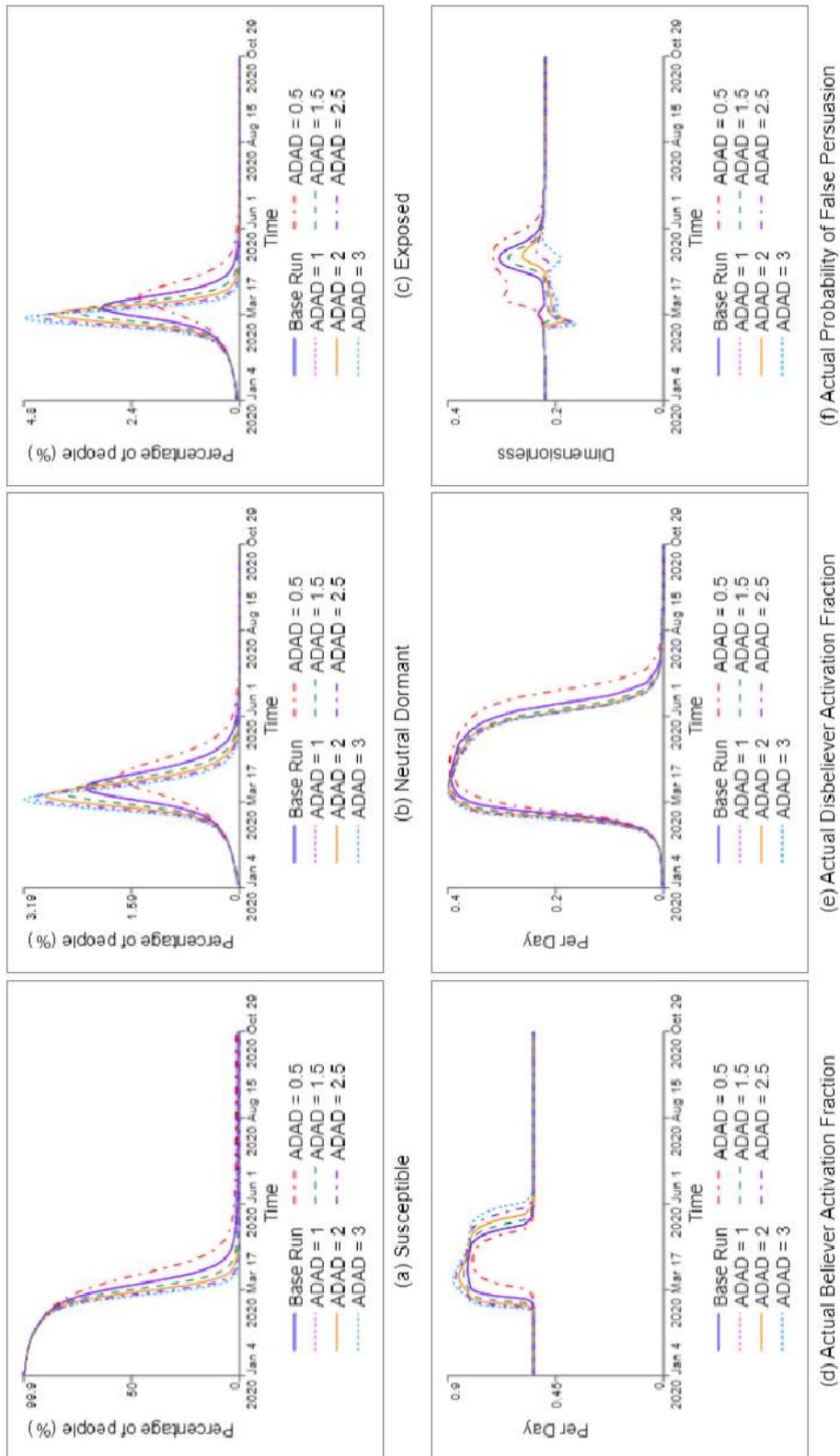


Figure C.14. Sensitivity to *Average Disbeliever Active Duration* (Base = 1)
(Panel 2)

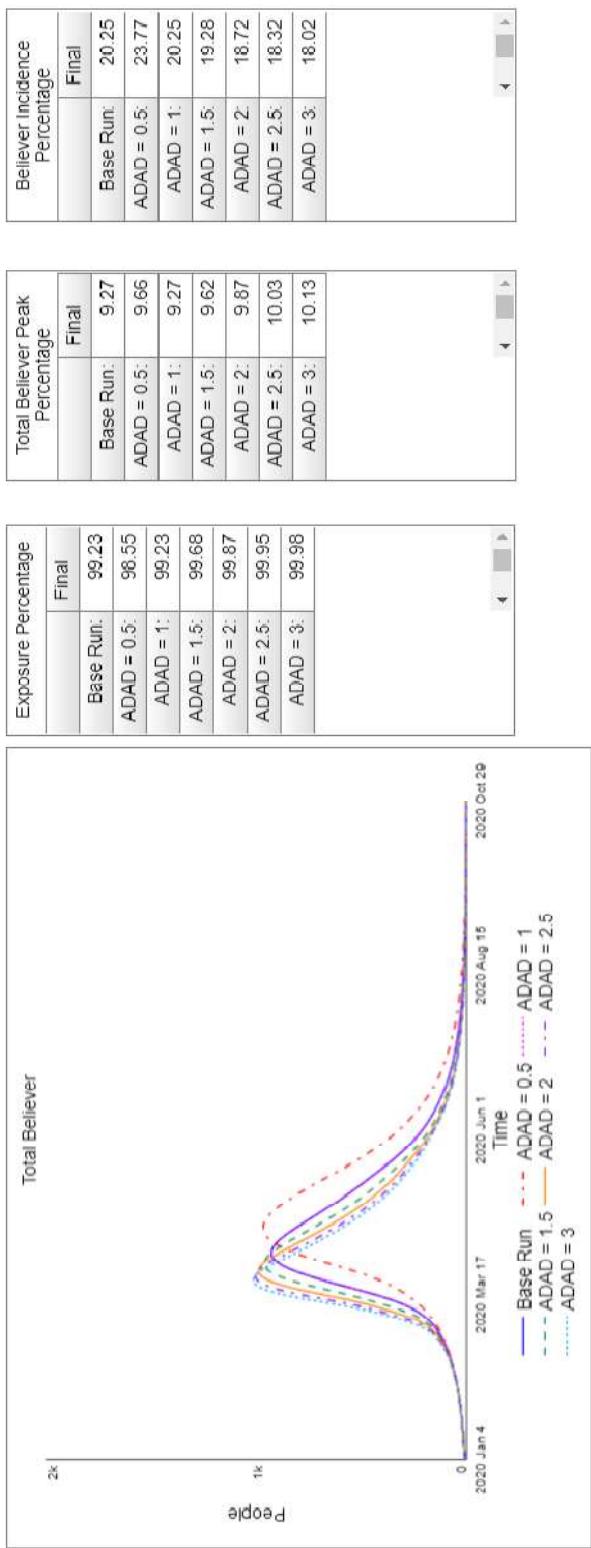


Figure C.15. Sensitivity to *Average Disbeliever Active Duration* (Base = 1)
(Panel 3)

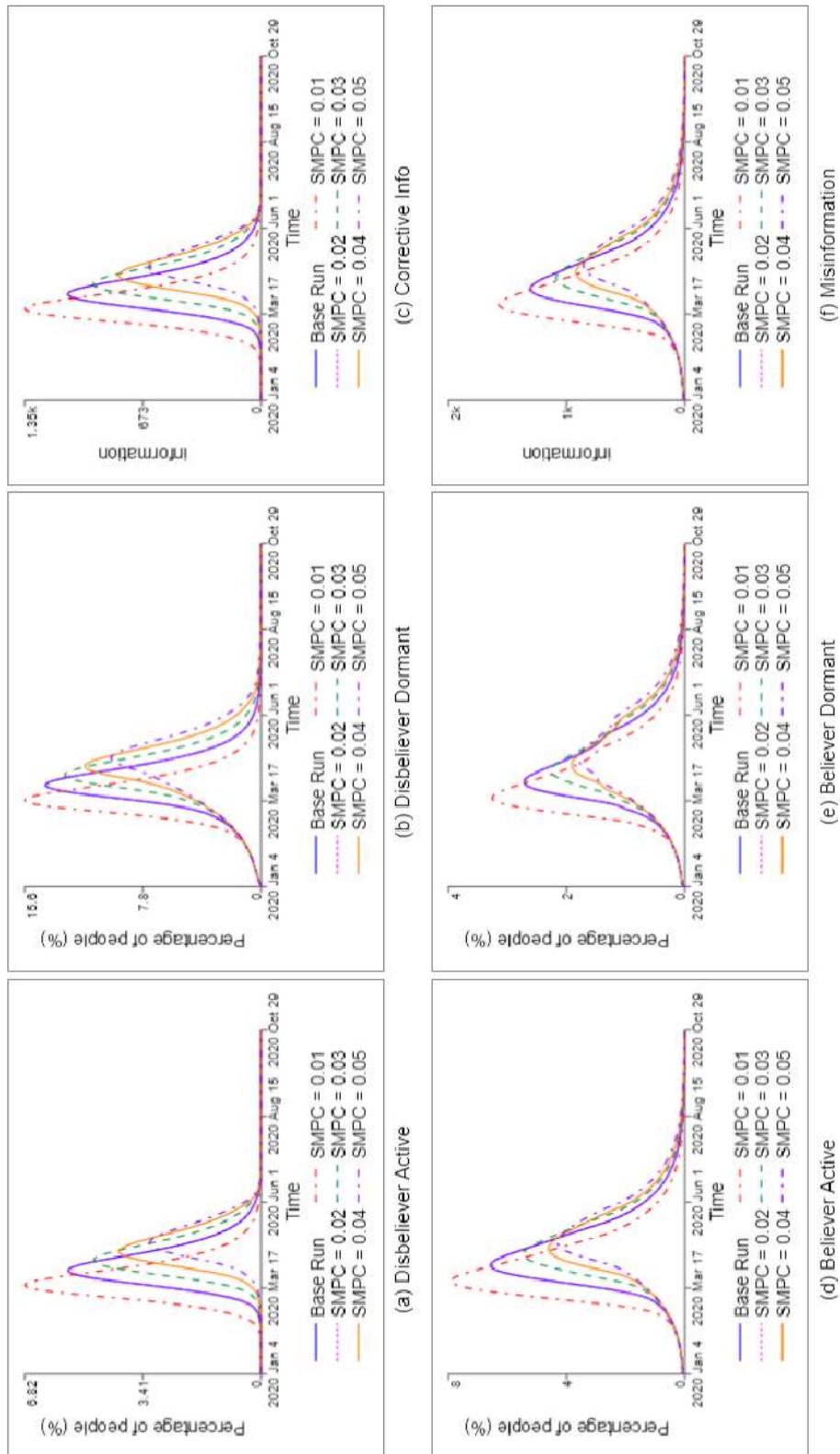


Figure C.16. Sensitivity to *Standard Misinformation per capita* (Base = 0.02)
(Panel 1)

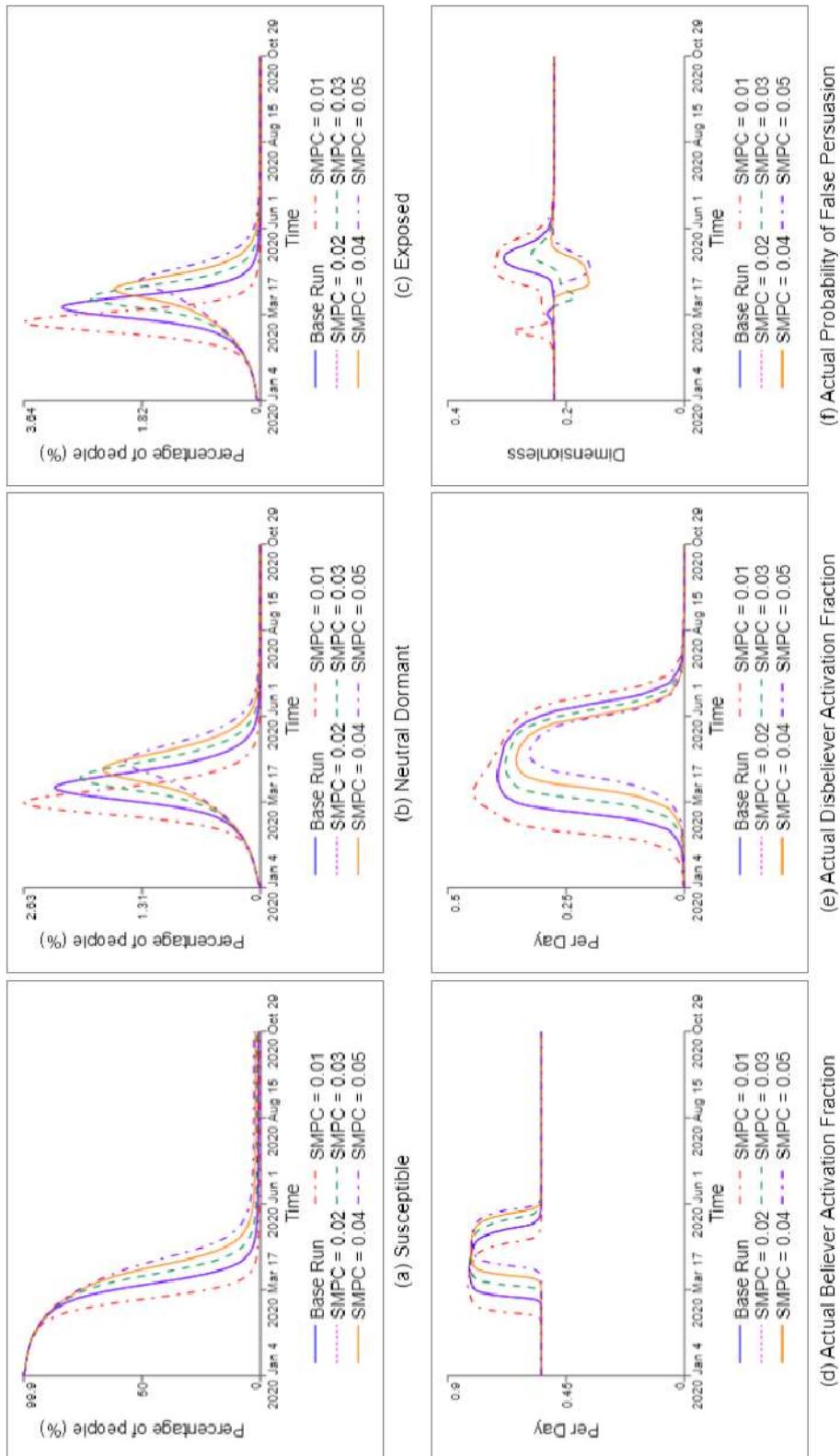


Figure C.17. Sensitivity to *Standard Misinformation per capita* (Base = 0.02)
(Panel 2)

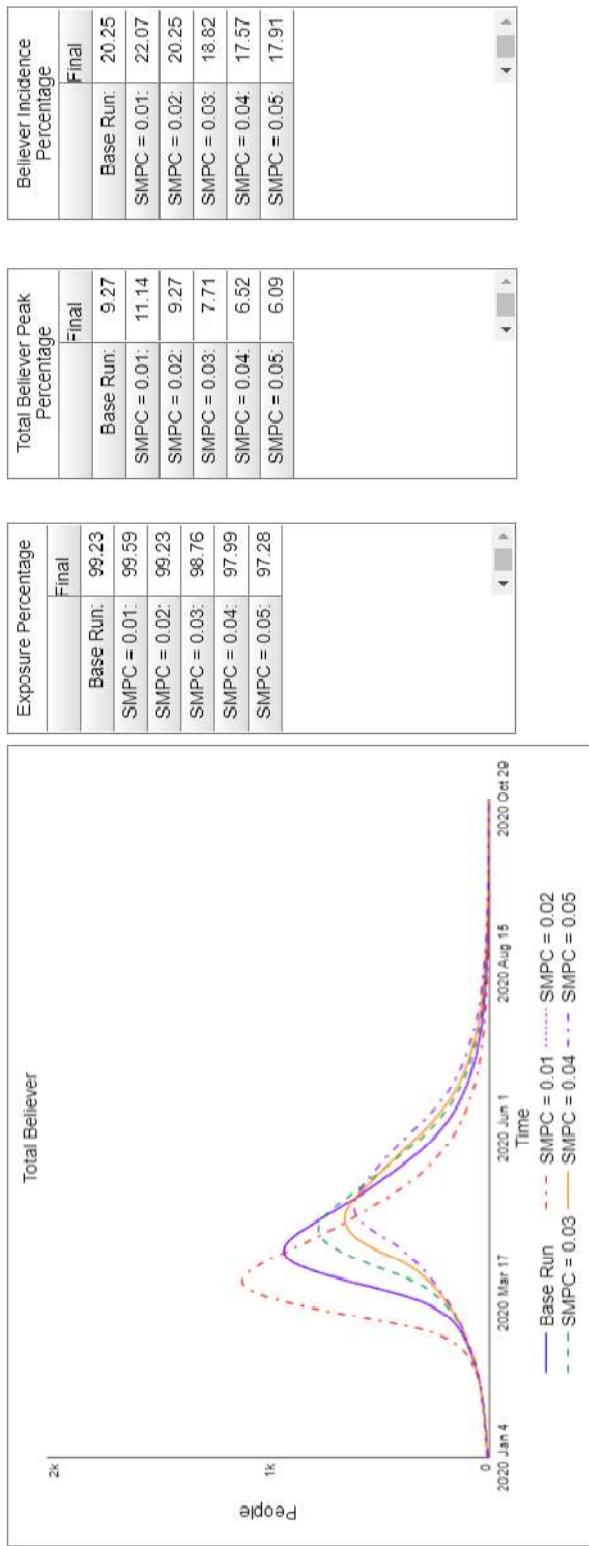


Figure C.18. Sensitivity to *Standard Misinformation per capita* (Base = 0.02)
 (Panel 3)

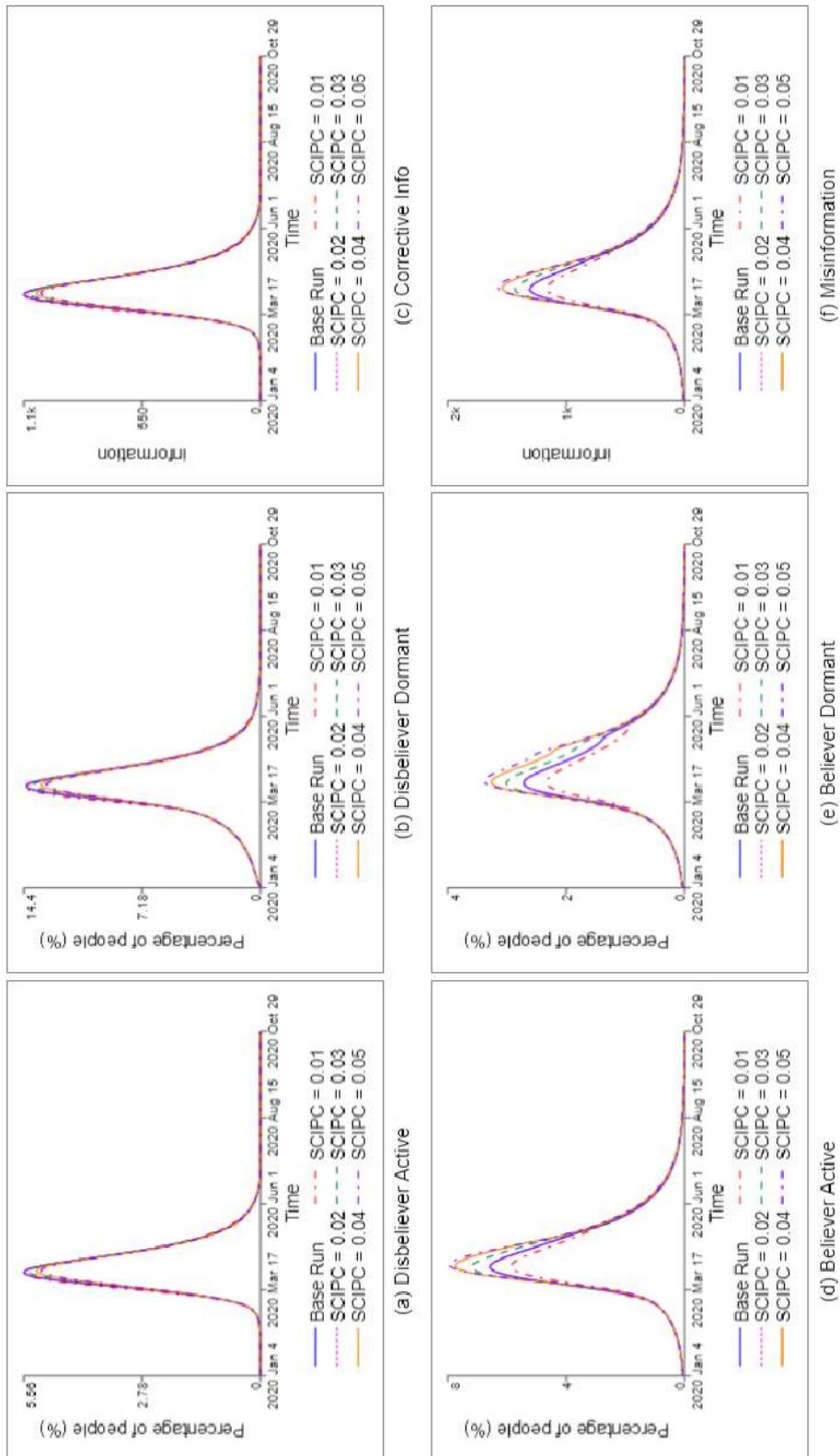


Figure C.19. Sensitivity to *Standard Corrective Info per capita* (Base = 0.02)
(Panel 1)

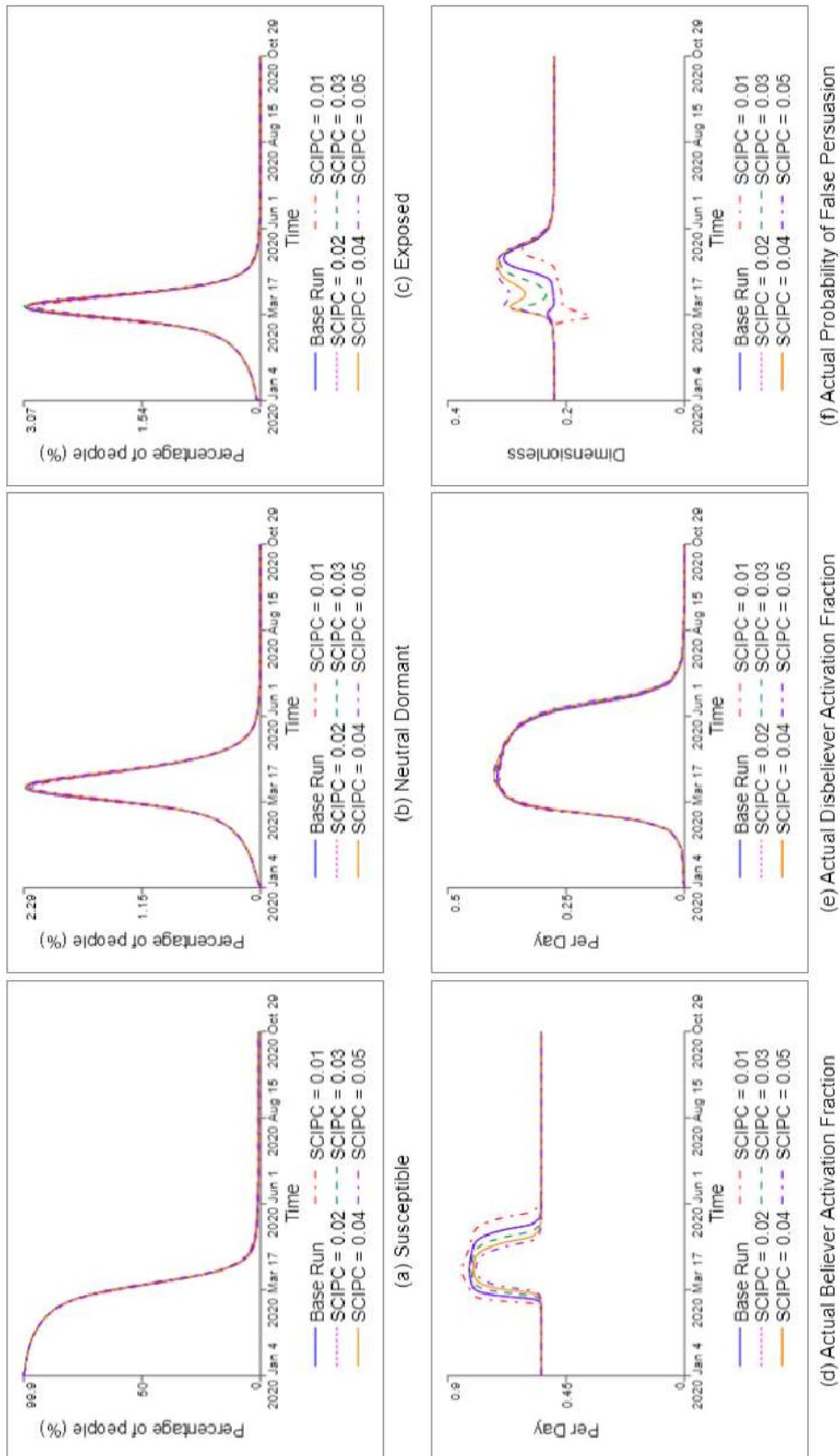


Figure C.20. Sensitivity to *Standard Corrective Info per capita* (Base = 0.02)
(Panel 2)

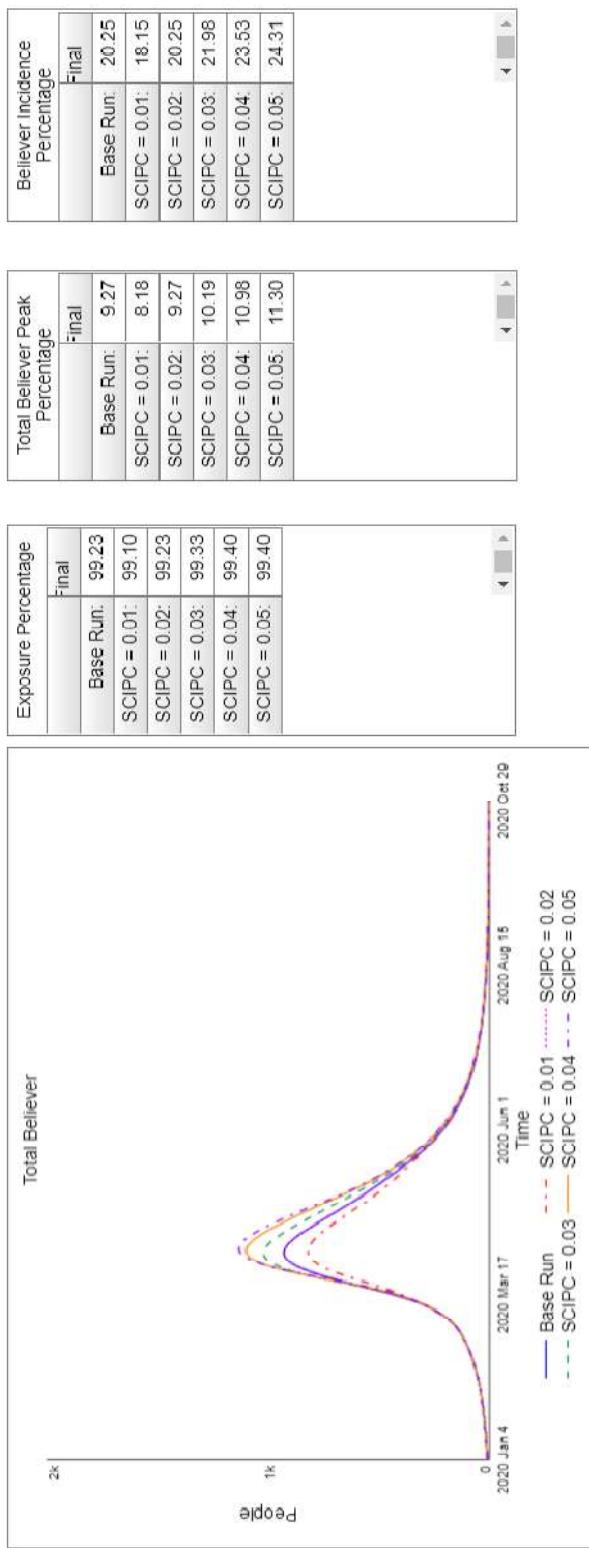


Figure C.21. Sensitivity to *Standard Corrective Info per capita* (Base = 0.02)
(Panel 3)

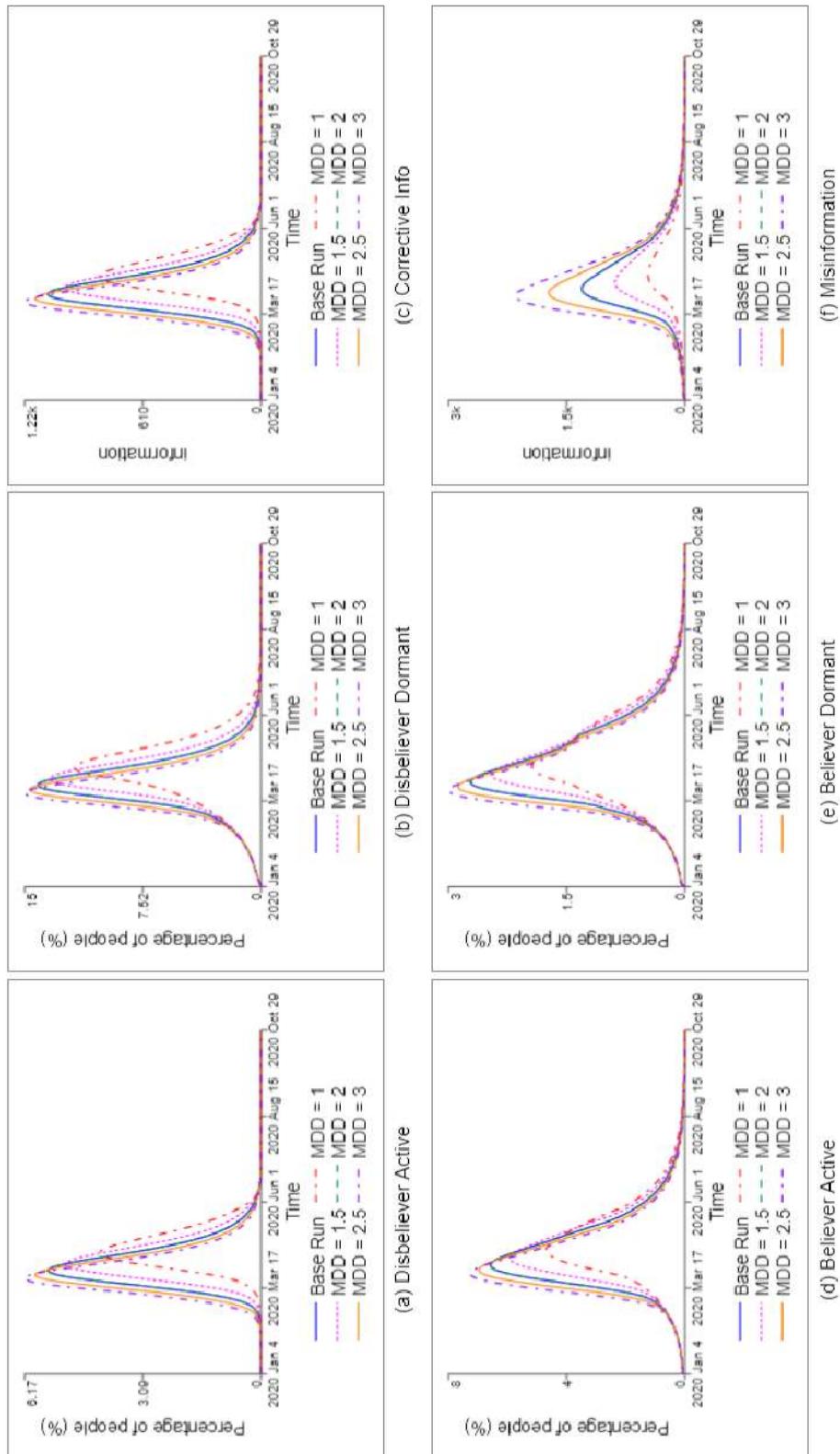


Figure C.22. Sensitivity to *Misinformation Depreciation Time* (Base = 2)
(Panel 1)

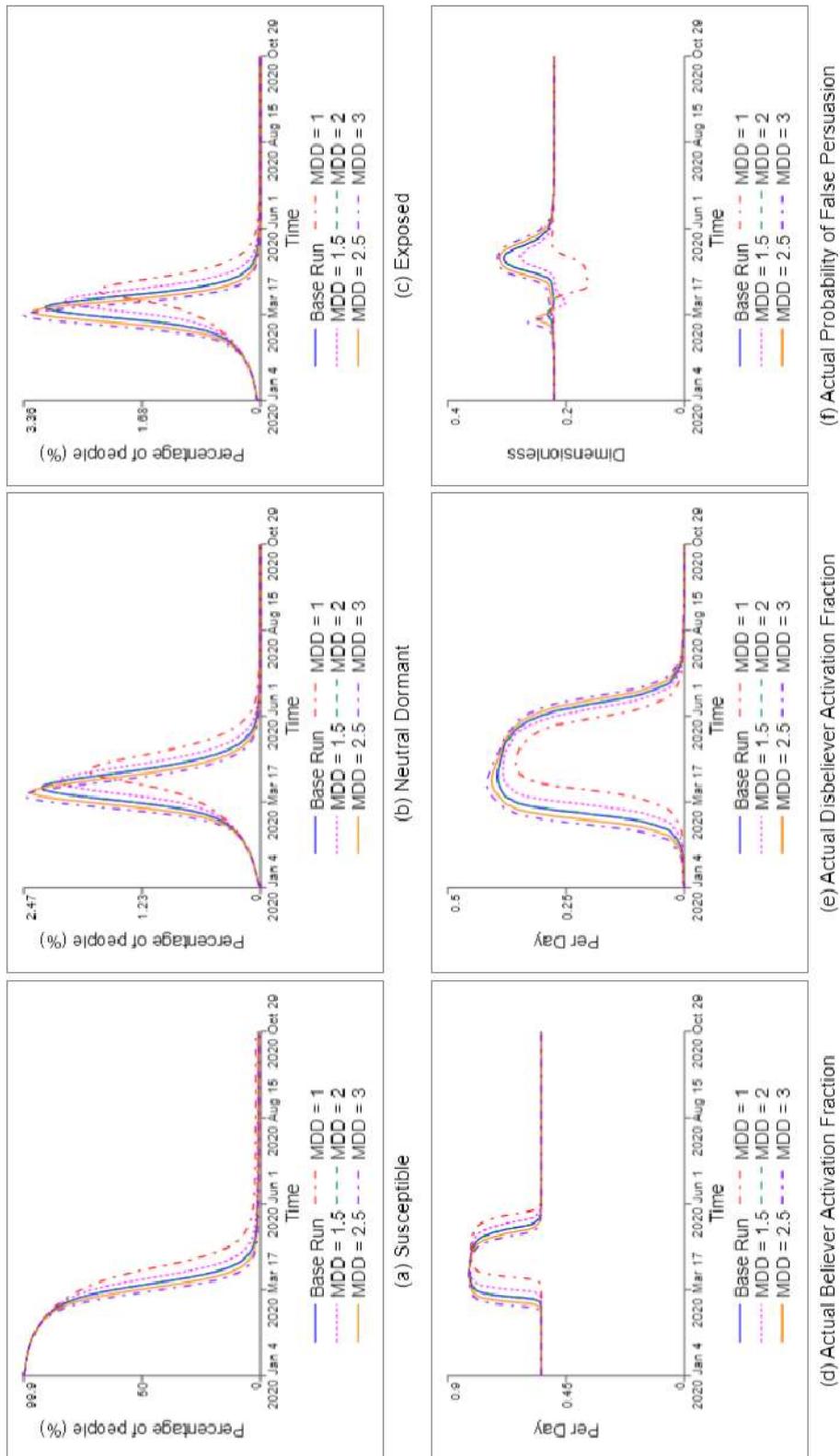


Figure C.23. Sensitivity to *Misinformation Depreciation Time* (Base = 2)
(Panel 2)

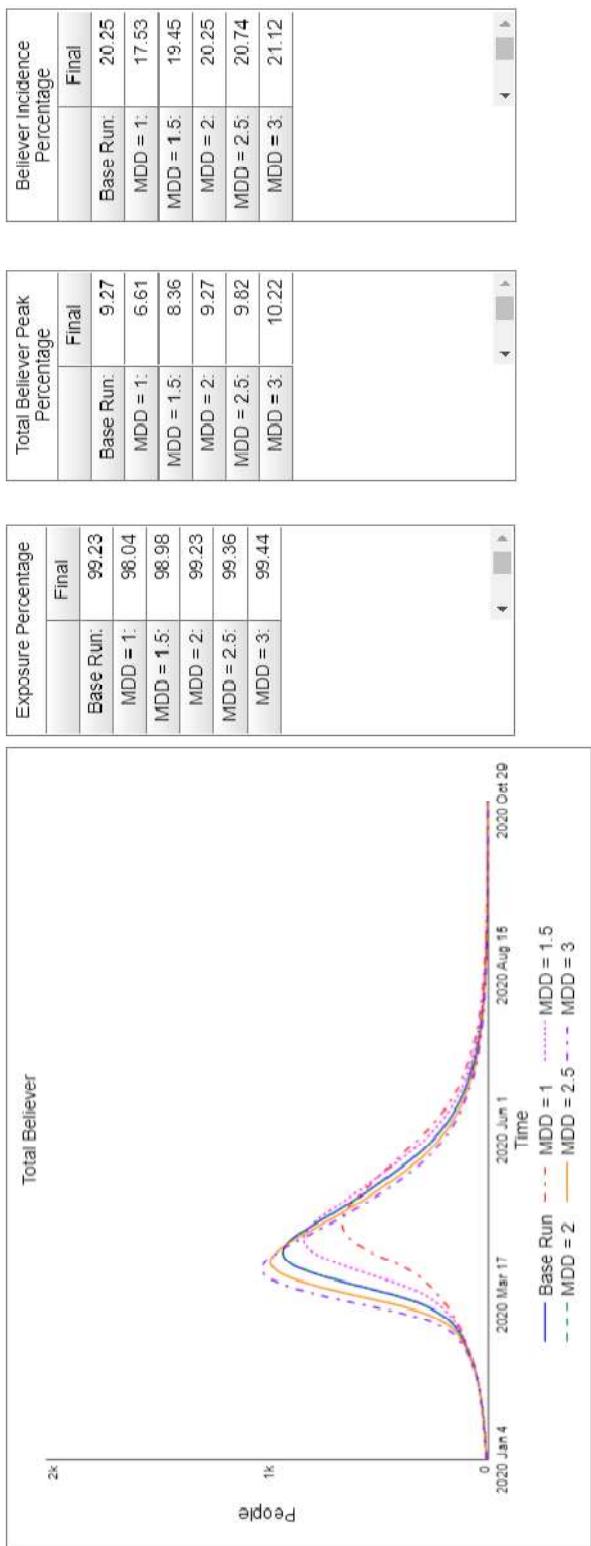


Figure C.24. Sensitivity to *Misinformation Depreciation Time* (Base = 2)
(Panel 3)

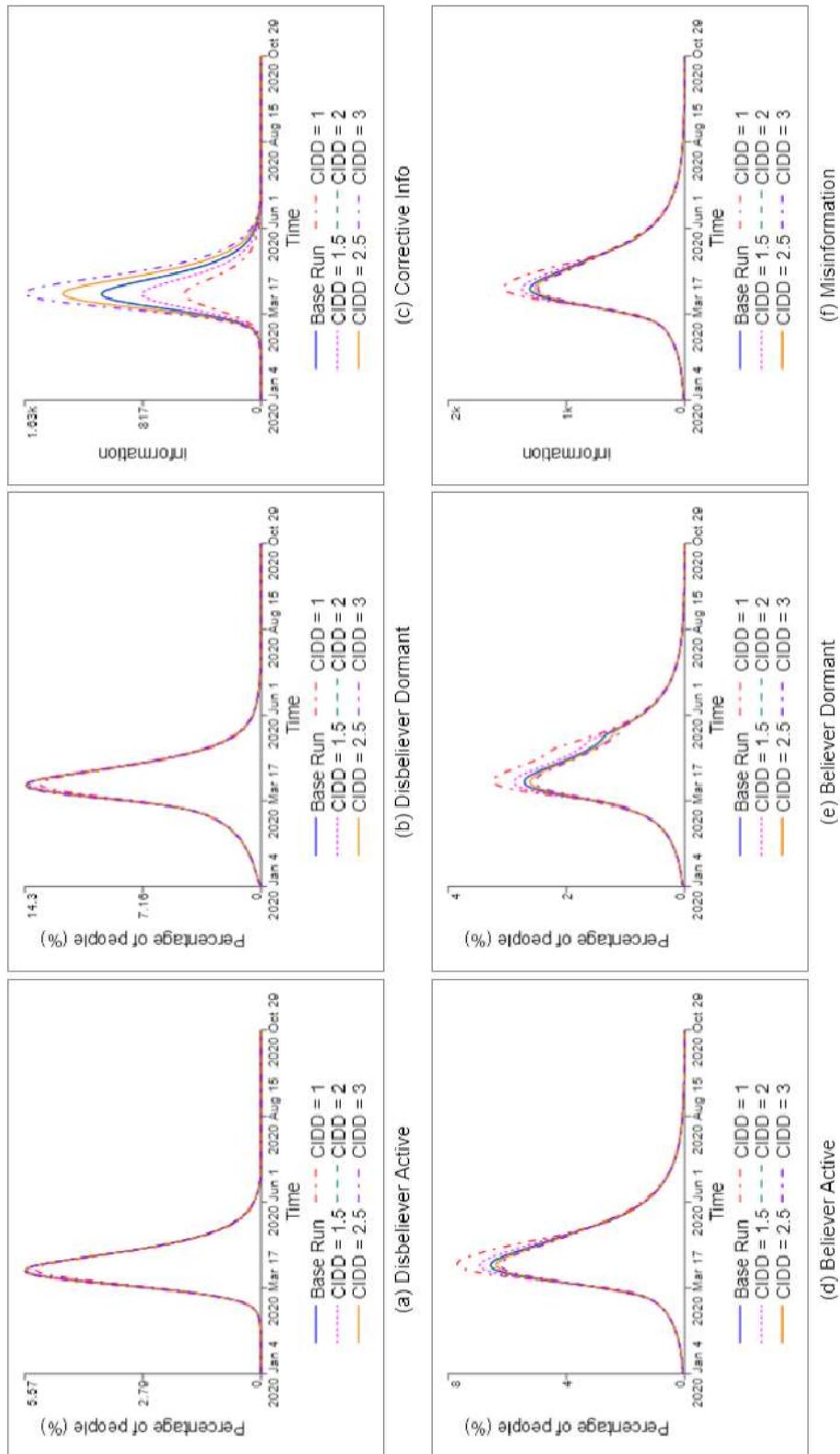


Figure C.25. Sensitivity to *Corrective Info Depreciation Time* (Base = 2)
(Panel 1)

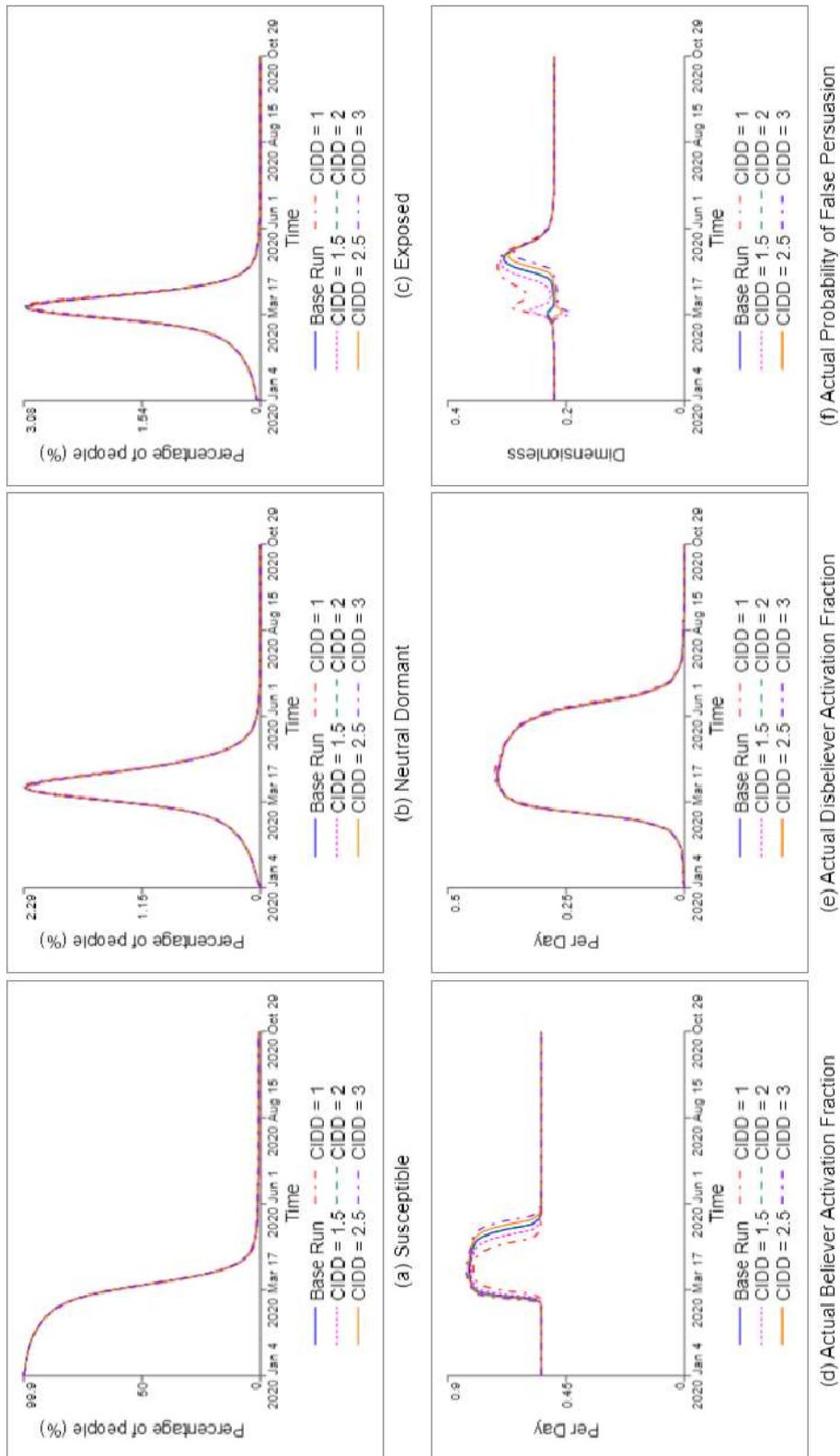


Figure C.26. Sensitivity to *Corrective Info Depreciation Time* (Base = 2)
(Panel 2)

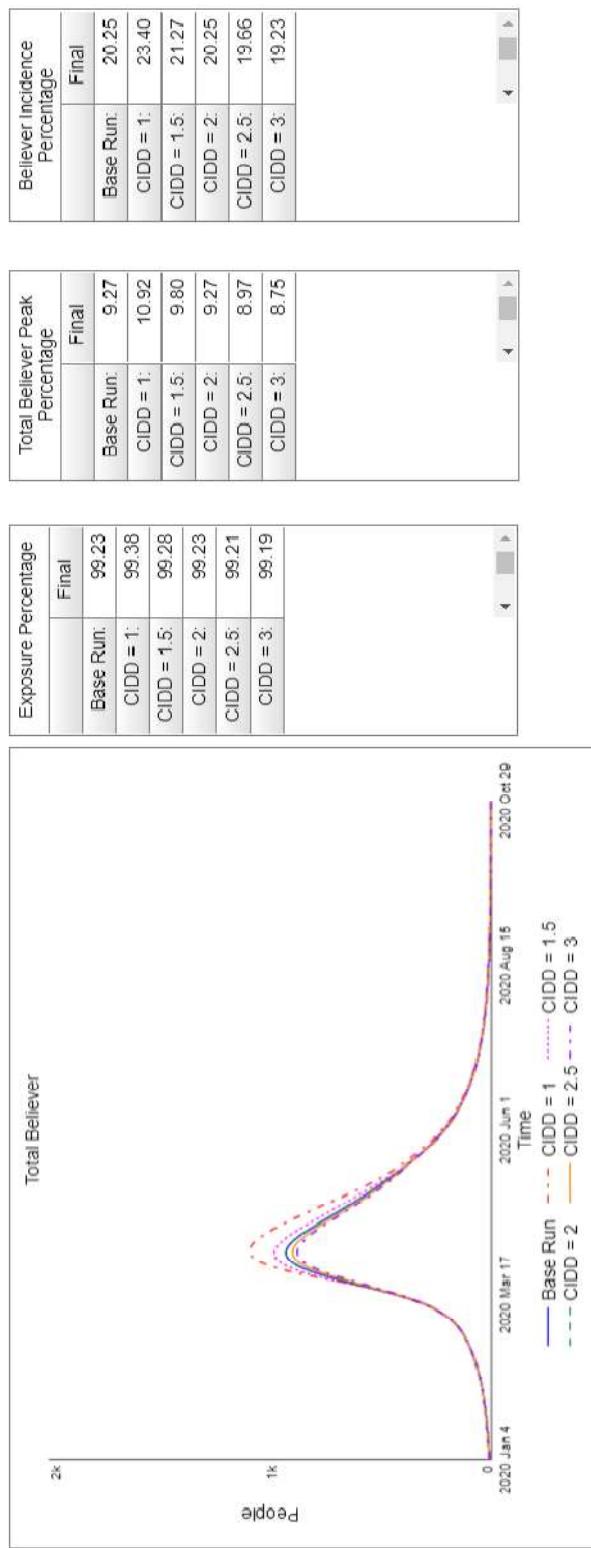


Figure C.27. Sensitivity to *Corrective Info Depreciation Time* (Base = 2)
(Panel 3)

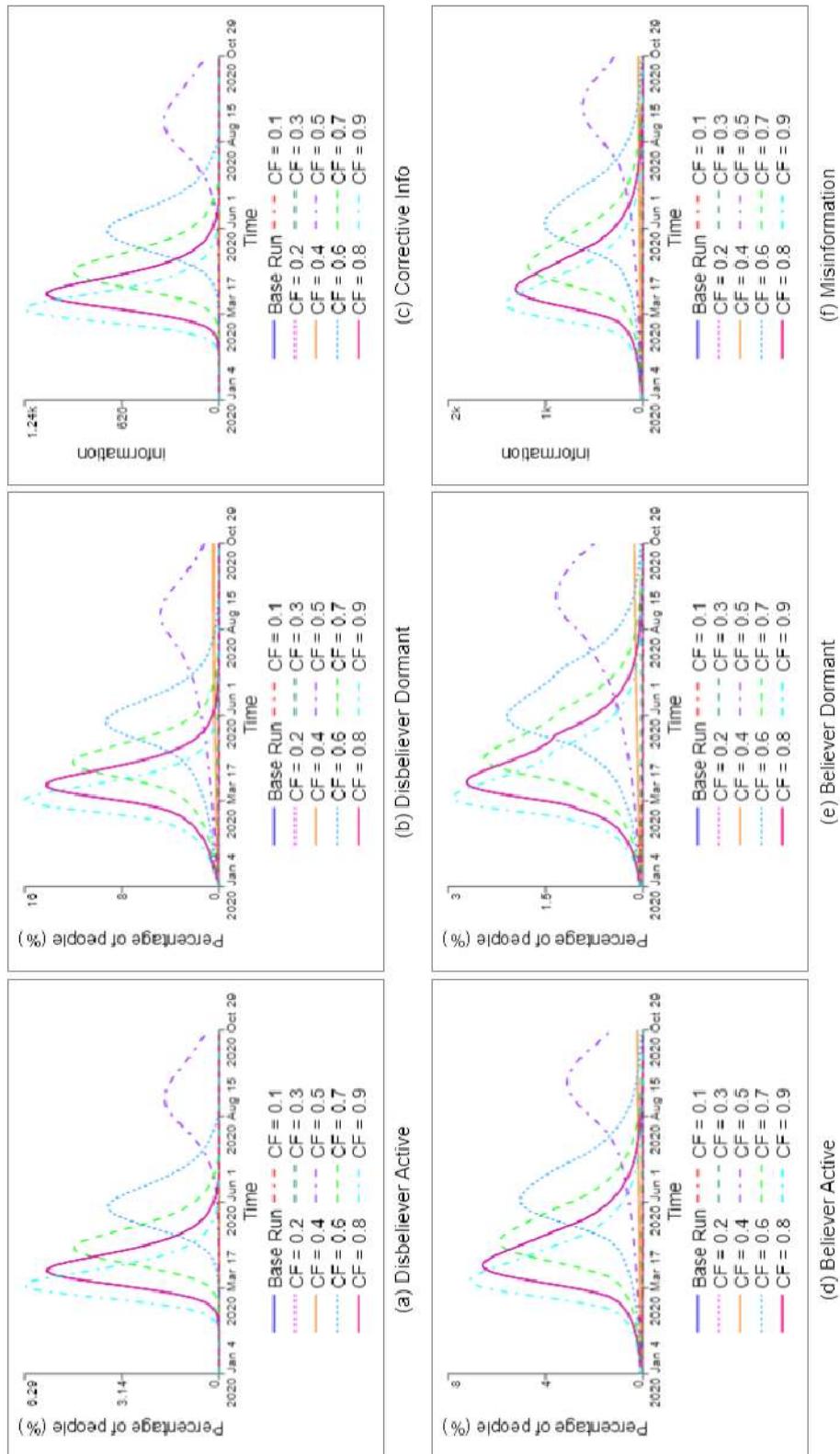


Figure C.28. Sensitivity to *Contact Fraction* (Base = 0.8)
(Panel 1)

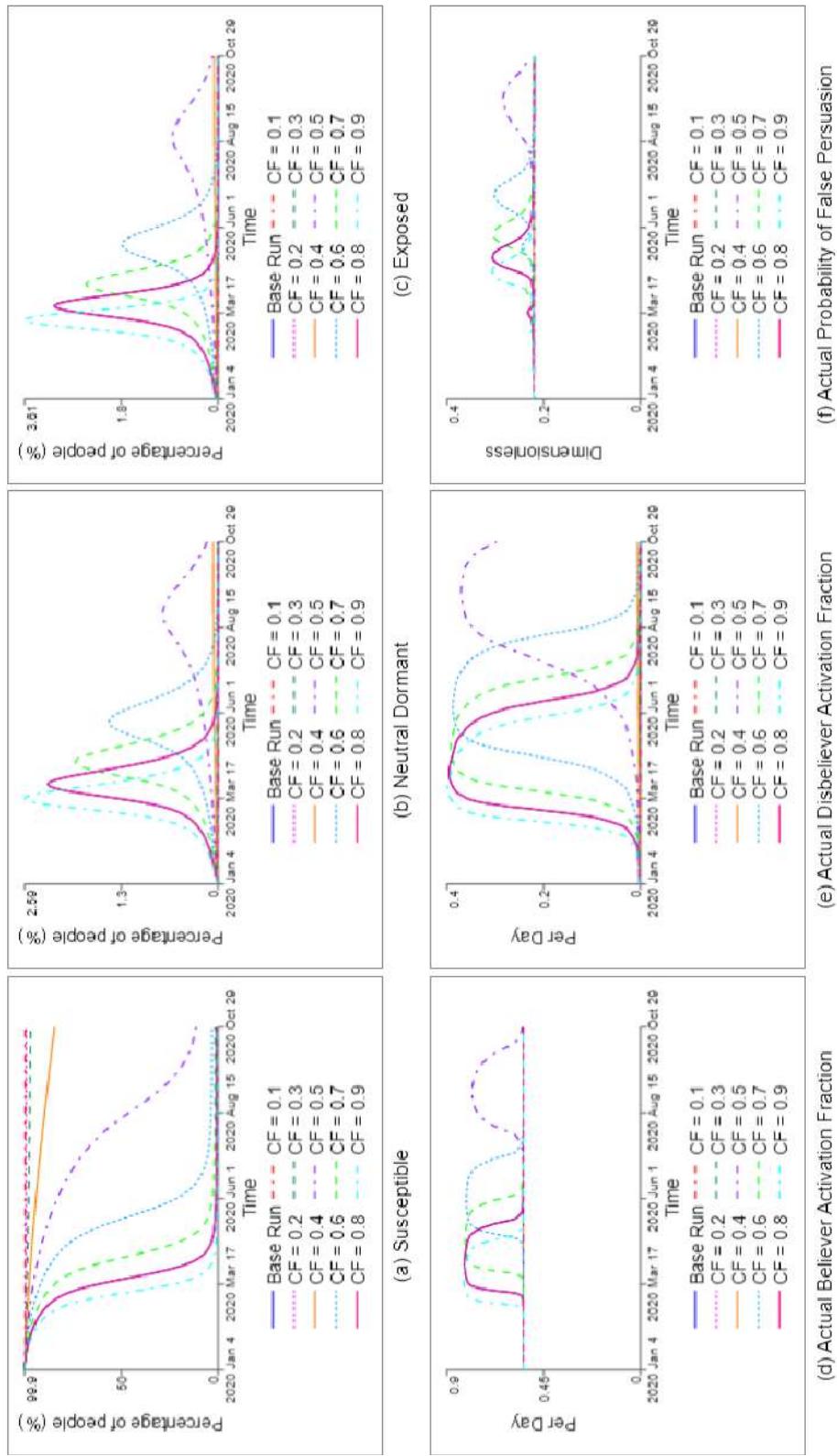


Figure C.29. Sensitivity to *Contact Fraction* (Base = 0.8)
(Panel 2)

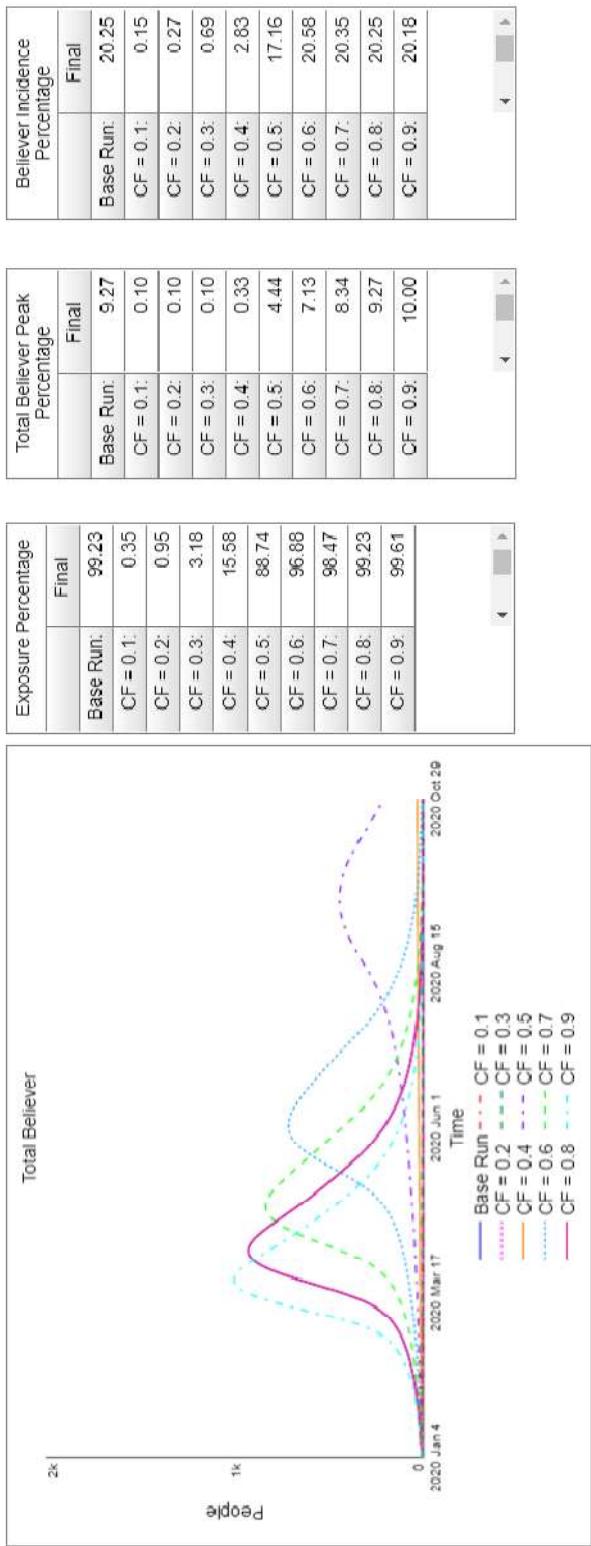


Figure C.30. Sensitivity to *Contact Fraction* (Base = 0.8)
(Panel 3)

REFERENCES

- Agley, J., & Xiao, Y. (2021). Misinformation about COVID-19: evidence for differential latent profiles and a strong association with trust in science. *BMC Public Health*, 21(1), 1–12. <https://doi.org/10.1186/s12889-020-10103-x>
- Ahmed, W., Vidal-Alaball, J., Downing, J., & Seguí, F. L. (2020). COVID-19 and the 5G conspiracy theory: Social network analysis of twitter data. *Journal of Medical Internet Research*, 22(5), 1–9. <https://doi.org/10.2196/19458>
- Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. 31(2), 211–236.
- Ammara, U., Bukhari, H., & Qadir, J. (2020). Analyzing Misinformation Through The Lens of Systems Thinking. Tto, 55–63.
- Atakan, D., & Şen, A. (Eds.). (2022, August 2). Turkish doctor at target of anti-vaxxers finds calf tongues in front of office. *Duvar English*. Retrieved August 8, 2022, from <https://www.duvarenglish.com/turkish-doctor-prof-esin-senol-at-target-of-anti-vaxxers-finds-calf-tongues-in-front-of-office-news-61091>.
- Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review*, 12(3), 183–210. [https://doi.org/10.1002/\(sici\)1099-1727\(199623\)12:3<183::aid-sdr103>3.0.co;2-4](https://doi.org/10.1002/(sici)1099-1727(199623)12:3<183::aid-sdr103>3.0.co;2-4)
- Barlas, Y., 2002, "System Dynamics: Systemic Feedback Modeling for Policy Analysis." in Knowledge for Sustainable Development - An Insight into the Encyclopedia of Life Support Systems, UNESCO-EOLSS Publishers, Paris, France; Oxford, UK, pp. 1131-1175.
- BBC. (2020, April 4). *Mast fire probe amid 5G coronavirus claims*. BBC News. Retrieved March 17, 2022, from <https://www.bbc.com/news/uk-england-52164358#>
- Brewis, H. (2020, April 14). *Nightingale Hospital Phone Mast attacked as 5G conspiracy theory rages*. London Evening Standard | Evening Standard. Retrieved March 17, 2022, from <https://www.standard.co.uk/news/uk/nhs-nightingale-phone-mast-arson-attack-5g-conspiracy-a4414351.html>
- Bridgman, A., Merkley, E., Zhilin, O., Loewen, P. J., Owen, T., & Ruths, D. (2021). Infodemic Pathways: Evaluating the Role That Traditional and Social Media Play in

Cross-National Information Transfer. *Frontiers in Political Science*, 3(March).
<https://doi.org/10.3389/fpos.2021.648646>

Bruns, A., Harrington, S., & Hurcombe, E. (2020). ‘Corona? 5G? or both?’: the dynamics of COVID-19/5G conspiracy theories on Facebook. *Media International Australia*, 177(1), 12–29. <https://doi.org/10.1177/1329878X20946113>

Caled, D., & Silva, M. J. (2021). Digital media and misinformation: An outlook on multidisciplinary strategies against manipulation. In *Journal of Computational Social Science* (Issue 0123456789). Springer Singapore.
<https://doi.org/10.1007/s42001-021-00118-8>

Chan, M. pui S., Jones, C. R., Hall Jamieson, K., & Albaracín, D. (2017). Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. *Psychological Science*, 28(11), 1531–1546.
<https://doi.org/10.1177/0956797617714579>

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The Truth About the Truth: A Meta-Analytic Review of the Truth Effect. *Personality and Social Psychology Review*, 14(2), 238–257. <https://doi.org/10.1177/1088868309352251>

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29. <https://doi.org/10.1038/s44159-021-00006-y>

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning & Verbal Behavior*, 16(1), 107–112. [https://doi.org/10.1016/S0022-5371\(77\)80012-1](https://doi.org/10.1016/S0022-5371(77)80012-1)

Kauk, J., Kreysa, H., & Schweinberger, S. R. (2021). Understanding and countering the spread of conspiracy theories in social networks: Evidence from epidemiological models of Twitter data. *PLOS ONE*, 16(8), e0256179.
<https://doi.org/10.1371/journal.pone.0256179>

Kumar, S., & Shah, N. (2018). False Information on Web and Social Media: A Survey. ArXiv, April.

Langguth, J., Filkuková, P., Brenner, S., Schroeder, D. T., & Pogorelov, K. (2022). COVID-19 and 5G conspiracy theories: long term observation of a digital wildfire. *International Journal of Data Science and Analytics*.
<https://doi.org/10.1007/s41060-022-00322-3>

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096.
<https://doi.org/10.1126/science.aao2998>

Lee Howell et al. Digital wildfires in a hyperconnected world. *WEFReport*, 3:15–94, 2013.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest, Supplement*, 13(3), 106–131.
<https://doi.org/10.1177/1529100612451018>

Lotito, Q. F., Zanella, D., & Casari, P. (2021). Realistic aspects of simulation models for fake news epidemics over social networks. *Future Internet*, 13(3).
<https://doi.org/10.3390/fi13030076>

Ma, S., & Zhang, H. (2021). Opinion Expression Dynamics in Social Media Chat Groups: An Integrated Quasi-Experimental and Agent-Based Model Approach. *Complexity*, 2021. <https://doi.org/10.1155/2021/2304754>

Moon, Angela. “Two-Thirds of American Adults Get News from Social Media: Survey.” Reuters, Thomson Reuters, 8 Sept. 2017, www.reuters.com/article/us-usa-internet-socialmedia/two-thirds-of-american-adults-get-news-from-social-media-survey-idUSKCN1BJ2A8.

O. Varol et al., in Proceedings of the 11th AAAI Conference on Web and Social Media (Association for the Advancement of Artificial Intelligence, Montreal, 2017), pp. 280–289.

Oyserman, D., & Dawson, A. (2020). Your Fake News, Our Facts. In *The Psychology of Fake News* (pp. 173–195). Routledge. <https://doi.org/10.4324/9780429295379-13>

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*, 31(7), 770–780.
<https://doi.org/10.1177/0956797620939054>

van Prooijen, J.-W., & Douglas, K. M. (2018). Belief in conspiracy theories: Basic principles of an emerging research domain. *European Journal of Social Psychology*, 48(7), 897–908. <https://doi.org/10.1002/ejsp.2530>

Sterman, J.D. *Business Dynamics: Systems Thinking and Modeling in a Complex World*. McGraw-Hill, Boston, 2000.

Trevors, G., & Duffy, M. C. (2020). Correcting COVID-19 Misconceptions Requires Caution. *Educational Researcher*, 49(7), 538–542.
<https://doi.org/10.3102/0013189X20953825>

Wu, L., Morstatter, F., Carley, K. M., & Liu, H. (2019). Misinformation in social media: Definition, manipulation, and detection. *ACM SIGKDD Explorations Newsletter*, 21(2), 80–90.
https://www.public.asu.edu/~huanliu/papers/Misinformation_LiangWu2019.pdf

Vicario, M. Del, Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences of the United States of America*, 113(3), 554–559. <https://doi.org/10.1073/pnas.1517441113>

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

Zhao, Z., Zhao, J., Sano, Y., Levy, O., Takayasu, H., Takayasu, M., Li, D., Wu, J., & Havlin, S. (2020). Fake news propagates differently from real news even at early stages of spreading. *EPJ Data Science*, 9(1). <https://doi.org/10.1140/epjds/s13688-020-00224-z>

Zhou, X., & Zafarani, R. (2018). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. <https://doi.org/10.1145/3395046>