

## Misinformation can prevent the suppression of epidemics

Andrei Sontag<sup>†</sup>, Tim Rogers<sup>†</sup>, Christian Yates<sup>†\*</sup>

<sup>†</sup>*Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, UK*

**ABSTRACT.** The effectiveness of non-pharmaceutical interventions, such as mask-wearing and social distancing, as control measures for pandemic disease relies upon a conscientious and well-informed public who are aware of and prepared to follow advice. Unfortunately, public health messages can be undermined by competing misinformation and conspiracy theories, spread virally through communities that are already distrustful of expert opinion. In this article, we propose and analyse a simple model of the interaction between disease spread and (mis-)information dynamics in a heterogeneous population composed of both trusting individuals who seek quality information and will take precautionary measures, and distrusting individuals who are susceptible to misinformation. We show that, as the density of the distrusting population increases, the model passes through a phase transition to a state in which major outbreaks cannot be suppressed. Our work highlights the urgent need for effective measures to combat the spread of misinformation.

**Keywords:** Epidemic modelling, information spread

### 1. INTRODUCTION

Non-pharmaceutical interventions, such as mask-wearing and social distancing, are important tools to tackle the spread of emergent pandemic disease [1]. They can be used to slow the progress of an outbreak or suppress it completely if carefully implemented and strictly adhered to [2, 3]. There are, however, substantial drawbacks, including immediate economic damage to certain industries and longer-term negative effects of social isolation on the wellbeing of individuals. Beyond any rational cost-benefit analysis, these measures also evoke a strong emotional response. The concept of a sudden government demand for people to change their behaviour and limit social interaction in the face of an unseen enemy is at odds with prevailing western liberal ideologies of the importance of individual freedoms. Unsurprisingly, these policies do not enjoy universal support and public perceptions are a key consideration in deciding implementation details – it has been widely reported that the UK government delayed action in spring 2020 for fear that “behavioural fatigue” would limit public adherence to social distancing rules [4].

Much of the effectiveness of social distancing strategies relies upon the behaviours of individuals; how conscientiously they follow advice and how successfully they encourage others to do the same. Many factors affect the degree to which individuals will change their behaviours: perception of the risk to themselves and to others; quality and availability of information; trust in the government and/or scientific establishment; the attitudes of their social contacts to these issues. Importantly, there is considerable feedback between these factors and the progression of the epidemic itself [5]. Previous modelling studies have examined the interplay between risk perception (modelled as an information spreading process), behaviour, and epidemic dynamics. Methodologies applied include: game theory models for social distancing [6]; opinion dynamics for behavioural change [7]; models of fear of infection and fear of the control [8, 9]; SIS-like models with aware categories on networks where behaviour changes according to health status and risk perception [10]; SIR models with aware and unaware states [11]; and awareness-driven reduction in contacts [12].

Of particular interest is the work by Funk et al. [13] investigating the interplay between behavioural change and disease spread through a stochastic SIR-like model [14] coupled with an awareness model in which individuals transmit information through a contact-based network. The authors characterise the dynamics of information spread by the following two event types: (1) *transmission of awareness*, occurring when an individual encounters someone with more up-to-date information, and (2) *fading of awareness*, which describes gradual relaxation back to a low-awareness state. The feedback between disease and information dynamics occurs through the generation of new information by infected individuals when they realise their condition, after which they and their contacts may adopt safer behaviours to limit further spread of the infection. In [15] (and similarly in independent work of Kiss et al. [16]) a simplified

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

\*c.yates@bath.ac.uk

version of this model was explored with only two states of aware-unaware individuals, in which it was shown that awareness-driven behaviour change can alter the outbreak conditions and can make it impossible for a disease to establish itself in the population.

The works discussed above share common (often unstated) assumptions about the populations they describe. They model homogeneous populations of individuals that believe in the existence of the disease, are prepared to alter their behaviour to reduce transmission, and have been correctly informed how to do so. Recent evidence sadly shows the extent to which these assumptions fail [17–21].

In this study, we extend the established model of Funk et al. [13] to include a subset of individuals who are distrustful of official advice, susceptible to misinformation, and more likely to engage in risky behaviours. We show that the presence of such dynamics can radically reduce the effectiveness of epidemic intervention strategies based on behaviour change. We observe a new dynamical threshold corresponding to a phase transition between regimes of successful disease suppression and large outbreaks. A theoretical estimate of the critical parameters for this transition is obtained in the limit of fast information spread.

The article is organised as follows. In Sec. 2, we propose a model of epidemics with behaviour change feedback in a heterogeneous population composed of trusting and distrusting individuals. Results from the modelling approach are presented in Sec. 3. We discuss the observed phase transition in Sec. 3.1 and derive a theoretical estimate for the fast spread of information limit in Sec. 3.2. We conclude with a brief discussion of our results in Sec. 4.

## 2. THE MODEL

We extend and modify the model presented in [13] with the introduction of a heterogeneous population composed of (i) *trusting* individuals who seek information of higher quality and will take precautionary measures, and (ii) *distrusting* individuals, who are susceptible to misinformation. We consider a population of fixed size  $N$  and SIR dynamics for disease spread [22]. In the SIR model, the population is compartmentalised into three disease states: susceptible  $S$ , infected  $I$  and recovered  $R$ . We distinguish subpopulations in our model by two indexes giving the behavioural group to which they belong, i.e., trusting (T) or distrusting (D), and the quality,  $i$ , of the information they possess. Information quality decreases with increasing index  $i$  so that, for example, an individual in the group  $S_{T,0}$  is a susceptible member of the trusting population and possesses the best quality information.

The two groups of trusting and distrusting individuals are distinguished by the way they obtain information. Trusting individuals of any disease state  $X \in \{S, I, R\}$  with information of quality  $j$ ,  $X_{T,j}$ , will accept new information when they encounter individuals of any group with information of better quality  $i$ , i.e.,  $X_{T,j} + X_{Y,i} \rightarrow X_{T,i+1} + X_{Y,i}$ , where  $i < j$  and  $Y$  can be either trusting (T) or distrusting (D). The “+1” in the information subscript represents a loss in quality when the information is exchanged. Hence, trusting individuals possess the same information dynamics as assigned to all individuals in [13]. In contrast, distrusting individuals  $X_{D,j}$  receive information when encountering individuals with lower quality information, i.e.,  $X_{D,j} + X_{Y,i} \rightarrow X_{D,i+1} + X_{Y,i}$ , where  $i > j$ . Parameters  $\alpha_T$  and  $\alpha_D$  control the encounter rates for each subpopulation.

Besides information transmission, we also assume a *fading* effect. Due to the fading effect, information loses quality if not refreshed, increasing the individual’s index by 1, i.e.,  $X_i \rightarrow X_{i+1}$ . This occurs at a rate  $\lambda$ . This dynamic accounts for reluctance of the population to continue with inconvenient measures when they do not perceive immediate risk.

The described information dynamics lead to the following equations representing the changes of population size at each information level  $k$  for each group of trusting and distrusting individuals,

$$\frac{dX_{T,k}}{dt} = -\frac{\alpha_T}{N} X_{T,k} \sum_{i=0}^{k-2} (X_{T,i} + X_{D,i}) + \frac{\alpha_T}{N} (X_{T,k-1} + X_{D,k-1}) \sum_{i=k+1}^{\infty} X_{T,i} - \lambda X_{T,k} + \lambda X_{T,k-1}, \quad (1)$$

$$\frac{dX_{D,k}}{dt} = -\frac{\alpha_D}{N} X_{D,k} \sum_{i=k+1}^{\infty} (X_{T,i} + X_{D,i}) + \frac{\alpha_D}{N} (X_{T,k-1} + X_{D,k-1}) \sum_{i=0}^{k-2} X_{D,i} - \lambda X_{D,k} + \lambda X_{D,k-1}, \quad (2)$$

which are valid for  $k \in \{0, 1, \dots\}$  if we take  $X_{Y,j} = 0$  for  $j < 0$ . The first term in Eq. (1) corresponds to individuals at the information level  $k$  moving to a better quality of information after interacting with either trusting or distrusting better-informed individuals. The second term corresponds to trusting individuals with worse information that interact with individuals at the level  $k-1$ , acquiring information of quality  $k$ . The last two terms describe the changes due to the fading effect. Similarly, the first term in Eq. (2) gives the changes due to distrusting individuals with information quality  $k$  obtaining worse

quality of information. The second term is the gain due to individuals with better information quality receiving information from level  $k - 1$ . The final terms again describe the changes due to information fading. Once some level of disease awareness is present in the population, the changes in distribution of the population throughout the awareness levels will follow the dynamics given by Eqs. (1) and (2).

Denoting  $I_i = I_{T,i} + I_{D,i}$ ,  $t_i = S_{T,i} + I_{T,i} + R_{T,i}$ , and  $d_i = S_{D,i} + I_{D,i} + R_{D,i}$ , we obtain the following set of ordinary differential equations to describe the changes in the subpopulations due to disease and behaviour dynamics,

$$\begin{aligned} \frac{dS_{T,k}}{dt} = & -\frac{S_{T,k}}{N} \sum_{i=0}^{\infty} I_i \cdot (1 - \rho^i)(1 - \rho^k)\beta - \frac{\alpha_T}{N} S_{T,k} \sum_{i=0}^{k-2} (t_i + d_i) - \lambda S_{T,k} + \lambda S_{T,k-1} \\ & + \frac{\alpha_T}{N} (t_{k-1} + d_{k-1}) \sum_{i=k+1}^{\infty} S_{T,i}, \end{aligned} \quad (3)$$

$$\begin{aligned} \frac{dS_{D,k}}{dt} = & -\frac{S_{D,k}}{N} \sum_{i=0}^{\infty} I_i \cdot (1 - \rho^i)(1 - \rho^k)\beta - \frac{\alpha_D}{N} S_{D,k} \sum_{i=k+1}^{\infty} (t_i + d_i) - \lambda S_{D,k} + \lambda S_{D,k-1} \\ & + \frac{\alpha_D}{N} (t_{k-1} + d_{k-1}) \sum_{i=0}^{k-2} S_{D,i}, \end{aligned} \quad (4)$$

$$\frac{dI_{T,0}}{dt} = -\sigma I_{T,0} + \omega \sum_{i=1}^{\infty} I_{T,i} - \lambda I_{T,0}, \quad (5)$$

$$\begin{aligned} \frac{dI_{T,k}}{dt} = & \frac{S_{T,k}}{N} \sum_{i=0}^{\infty} I_i \cdot (1 - \rho^i)(1 - \rho^k)\beta - \sigma I_{T,k} - \frac{\alpha_T}{N} I_{T,k} \sum_{i=0}^{k-2} (t_i + d_i) - \lambda I_{T,k} + \lambda I_{T,k-1} \\ & + \frac{\alpha_T}{N} (t_{k-1} + d_{k-1}) \sum_{i=k+1}^{\infty} I_{T,i} - \omega I_{T,k}, \end{aligned} \quad (6)$$

$$\frac{dI_{D,0}}{dt} = -\sigma I_{D,0} + \omega \sum_{i=1}^{\infty} I_{D,i} - \frac{\alpha_D}{N} I_{D,0} \sum_{i=1}^{\infty} (t_i + d_i) - \lambda I_{D,0}, \quad (7)$$

$$\begin{aligned} \frac{dI_{D,k}}{dt} = & \frac{S_{D,k}}{N} \sum_{i=0}^{\infty} I_i \cdot (1 - \rho^i)(1 - \rho^k)\beta - \sigma I_{D,k} - \frac{\alpha_D}{N} I_{D,k} \sum_{i=k+1}^{\infty} (t_i + d_i) - \lambda I_{D,k} + \lambda I_{D,k-1} \\ & + \frac{\alpha_D}{N} (t_{k-1} + d_{k-1}) \sum_{i=0}^{k-2} I_{D,i} - \omega I_{D,k}, \end{aligned} \quad (8)$$

$$\frac{dR_{T,k}}{dt} = \sigma I_{T,k} - \frac{\alpha_T}{N} R_{T,k} \sum_{i=0}^{k-2} (t_i + d_i) - \lambda R_{T,k} + \lambda R_{T,k-1} + \frac{\alpha_T}{N} (t_{k-1} + d_{k-1}) \sum_{i=k+1}^{\infty} R_{T,i}, \quad (9)$$

$$\frac{dR_{D,k}}{dt} = \sigma I_{D,k} - \frac{\alpha_D}{N} R_{D,k} \sum_{i=k+1}^{\infty} (t_i + d_i) - \lambda R_{D,k} + \lambda R_{D,k-1} + \frac{\alpha_D}{N} (t_{k-1} + d_{k-1}) \sum_{i=0}^{k-2} R_{D,i}, \quad (10)$$

where  $\beta$  is the transmission rate in the absence of any protective measures,  $\sigma$  is the recovery rate, and  $\rho$  is a fixed parameter which measures the effectiveness of non-pharmaceutical interventions,  $0 < \rho < 1$ . The first terms in Eqs. (3), (4), (6), and (8) correspond to new infections of susceptible individuals. In contrast with the model [13], we assume that disease transmission is reduced by both infected and susceptible individuals reducing their contact rates, hence the factor  $(1 - \rho^i)(1 - \rho^k)$  in the transmission rates. In this case, infectious individuals at information level 0 do not infect any susceptibles during the period they are fully aware of their condition. This is the case for diseases for which self-isolation is imposed after the detection of the disease - assuming the disease is detected before the individual recovers.

The second link between disease and awareness dynamics is obtained through the second terms in Eqs. (5) and (7) that describe the information generation by infected individuals. Information is generated when infected individuals realise their condition, for example following a positive test result. This effect distinguishes between diseases with unmistakable symptoms and cases where the infection is contagious but asymptomatic, or where the infection might be mistaken for another condition. The parameter  $\omega$  then varies not only from illness to illness, but it also depends on the mechanisms to identify the disease quickly, such as fast testing and contact-tracing programmes.

Finally, the  $\sigma I_{Y,j}$  terms in Eqs.(5-10) correspond to the recovery of infected individuals. The remaining terms correspond to the previously described information dynamics (see Eqs.(1-2)), giving the changes

in awareness level due to information exchange and fading. Summing over the information subscripts leads to

$$\frac{dS}{dt} = -\frac{\tilde{\beta}SI}{N}, \quad \frac{dI}{dt} = \frac{\tilde{\beta}SI}{N} - \sigma I, \quad \frac{dR}{dt} = \sigma I, \quad (11)$$

where

$$\tilde{\beta}(\rho, \{S_j, I_i\}) = \beta \sum_{i,j=0}^{\infty} (1 - \rho^i)(1 - \rho^j) \frac{I_i S_j}{I S} \quad (12)$$

yields an effective transmission rate due to protective measures taken by informed individuals. Here,  $I_i = I_{C,i} + I_{D,i}$  and  $S_j = S_{C,j} + S_{D,j}$  as before, and  $I = \sum_i I_i$  and  $S = \sum_i S_i$ . This is similar to the SIR model, with  $\beta$  varying as the epidemic progresses. In fact, our model reduces to the SIR model if no awareness is present in the population and awareness generation is turned off, i.e., the population is distributed within subpopulations  $X_{Y,\infty}$ , and  $\omega = 0$ , respectively. In the SIR model, the epidemic threshold is at  $R_0 = \beta/\sigma = 1$  [23]. In our model, the effective reproduction number has a similar form,  $R_e = \tilde{\beta}/\sigma$ . If we start with a fully uninformed and susceptible population, in which case awareness arises only through the process of information generation, then  $\tilde{\beta}_{t=0} = \beta$ , and the epidemic threshold is again at  $R_e = \beta/\sigma$ . Only if a certain level of awareness were already present in the population at the beginning of the epidemic, would the threshold be reduced to  $R_e = \tilde{\beta}/\sigma$ , where  $\tilde{\beta}_{t=0} < \beta$ .

Notice that no transmission would occur if the population was completely informed with the best quality of information. This situation does not arise, however, because susceptibles can at best obtain an index 1 status if they are informed by infected individuals with first-hand information,  $I_0$ .

There is now a mutual feedback between information and disease dynamics, where information is produced by newly infected individuals, increasing the level of awareness in the trusting population and eliciting higher levels of protection through a reduction in contacts and adherence to other non-pharmaceutical interventions, hindering the spread of the disease. With sufficiently fast spread of information in the trusting population, leading to a quicker response to an increase in the number of infections, a fine balance can be achieved. The presence of distrusting individuals in the population reduces the effectiveness of the protection generated by the trusting population, threatening this delicate equilibrium. In the next section, we investigate the effects of varying the speed of information spread parameters,  $\alpha_T$  and  $\alpha_D$ , the information-mediated protection parameter,  $\rho$ , and the density of distrusting individuals in the population,  $d$ , on the course and outcome of an epidemic.

### 3. RESULTS

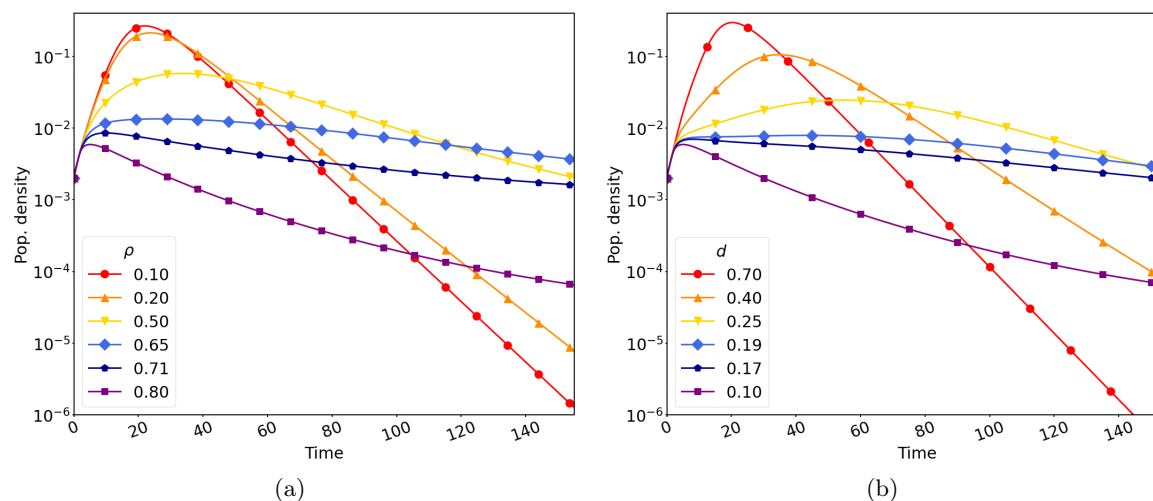
During an epidemic, the extent to which people comply with restrictions, such as mask-wearing, self-isolation, and social distancing, and the effectiveness of such protective measures is pivotal to control the spread of the disease [24]. In our model, the parameters that quantify these quantities are the density of distrusting individuals in the population,  $d$ , and the information-mediated protection parameter  $\rho$ .

#### 3.1. The effects of varying $\rho$ and $d$ .

Consider an initially unaware population composed of distrusting and trusting individuals with density  $d$  and  $1 - d$ , respectively, where a small number of infected individuals are introduced in the population. We assume that the composition of trusting and distrusting populations remains constant during the course of the epidemic. Initially, there is no awareness in the population, and consequently no protection stemming from behavioural change. Hence, the disease spreads following a classic *SIR* model, where the threshold  $R_0 = \frac{\beta}{\sigma} > 1$  gives the condition for an initial increase in the number of cases.

As in the SIR model, if  $R_0 < 1$ , the disease cannot spread in the population even in the absence of behavioural changes to control it. The awareness arising from the initial cases increases protection in the population and reduces the effective reproduction number  $R_e$  even further, speeding up the process of disease die out.

If  $R_0 > 1$ , after an initial phase of unrestricted spread, the infected individuals will start to become aware of their condition, generating fresh information in the population. For a single infected individual, the time taken is of order  $\omega^{-1}$ . This awareness then spreads according to the information dynamics, eliciting protective measures among the informed individuals and causing a reduction in the effective reproduction number  $R_e$ . If the response to new infections is sufficiently strong, a fine balance can be achieved, in which the increase in cases is counterbalanced by more cautious behaviour in the population. The presence of distrusting individuals, however, limits the effectiveness of the information-mediated



**Figure 1.** Stronger interventions (increasing  $\rho$ ) have a similar effect on disease dynamics as a more trusting population (decreasing  $d$ ). Curves show time series of the total number of infected  $I(t)$  for varying (a)  $\rho$  (fixed  $d = 0.1$ ) and (b)  $d$  (fixed  $\rho = 0.8$ ). In both cases  $\alpha_T = \alpha_D = 5$ ,  $\beta = 0.667$ ,  $\sigma = 0.133$  ( $R_e(0) = R_0 = 5$ ),  $\lambda = 0.2$ , and  $\omega = 0.333$ .

protection in the population, as they will only comply with protective measures if directly affected by the disease.

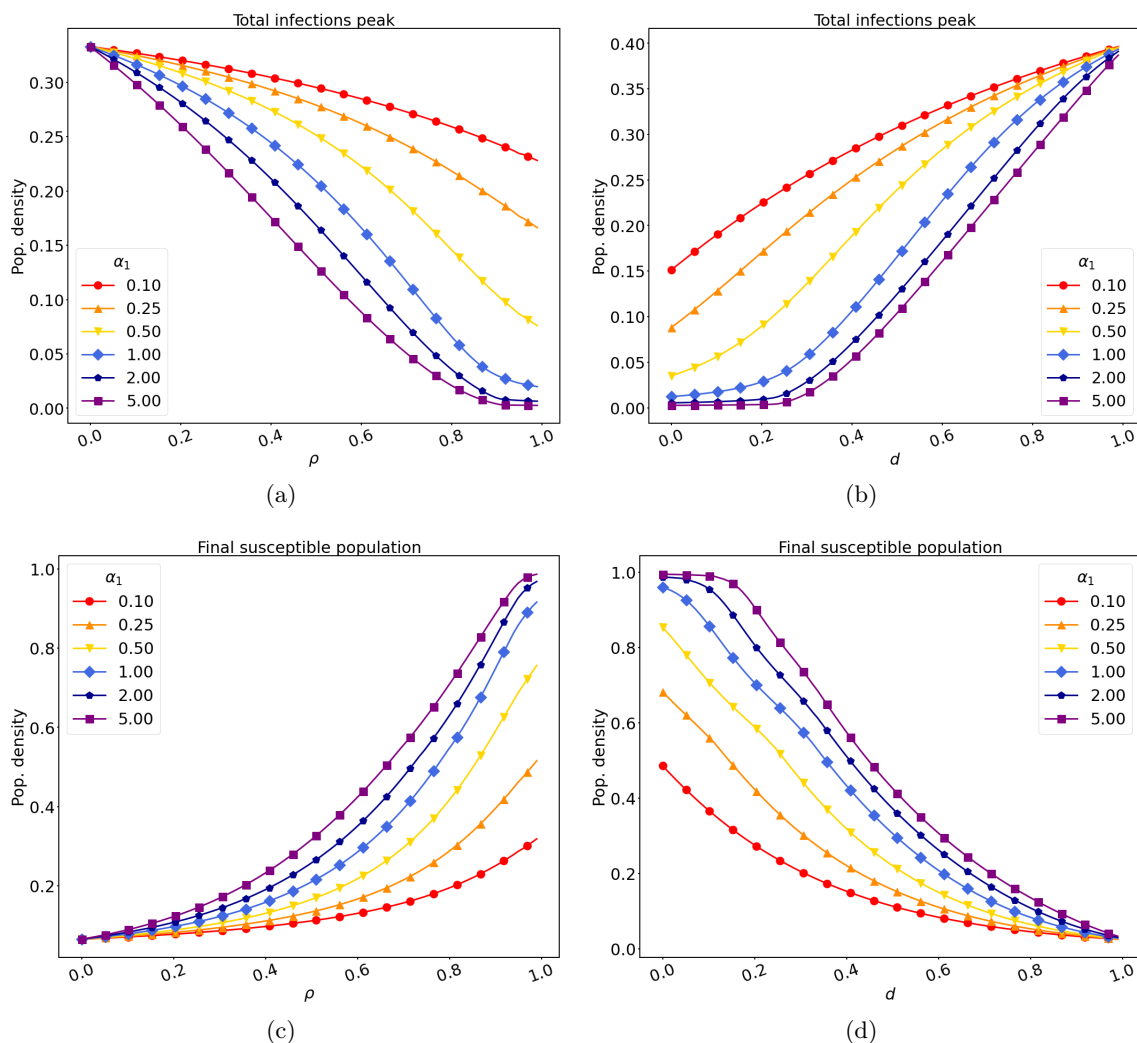
In Figure 1(a) we show the trajectories of the total number of infected over time,  $I(t)$ , for different  $\rho$  values, fixing the remaining parameters. Similarly, Figure 1(b) shows  $I(t)$  for different distrusting densities  $d$ . We have chosen  $R_0 > 1$  so that we see non-trivial infection dynamics. In both cases, we observe an initial increase in the number of infected individuals that is common to all curves. This corresponds to the unaware phase with  $R_0 > 1$ . After this initial phase of unrestricted contagion, there is a change in the curves' behaviour as the parameters vary. For small  $\rho$  and large  $d$  values, the elicited protection is not able to suppress the outbreak after the initial contagion phase, in which case we observe an effect similar to a reduction in  $\beta$  for the SIR model, where the peak of infection occurs later, and the epidemic duration extends. In contrast, for large  $\rho$  values and small densities  $d$ , the evoked protective measures taken by the informed individuals are sufficient to bring and keep  $R_e$  below 1. The epidemic outbreak is then suppressed, and we observe a peak of infection happening earlier in the epidemic, also reducing its duration.

For intermediate values, we observe the delicate balance between new infections, generating awareness, and the reduction in transmissibility by the informed individuals. This leads to a plateau in the number of infected individuals in the population and to really long epidemics. A similar effect has been observed by Weitz et al. [12] in the context of a SEIR model without awareness spread, but including awareness-driven behavioural changes. There are critical values  $\rho_c$  and  $d_c$  for which we observe a phase transition in the behaviour of the epidemic. For  $\rho < \rho_c$  (or  $d > d_c$ ), the epidemic spread is only mitigated, while for  $\rho > \rho_c$  ( $d < d_c$ ) the outbreak is suppressed by the behavioural changes in the population. The value of  $\rho_c$  depends on the density of distrusting individuals in the population, and similarly  $d_c$  depends on the information-mediated protection levels. We investigate this dependence,  $\rho_c = \rho_c(d)$ , further in Sec. 3.2.

In the real world, resources such as hospital staff, ventilators, and ICU beds are limited. Thus, knowing how many people will be affected at the same time and require medical assistance is essential to manage epidemics. Not only that, being aware of the factors that influence the size of the peak of infection is extremely valuable to aid policy-making. To explore further the effects of  $\rho$  and  $d$  on the peak of infection, we solved Eqs.(3-10) numerically for different values of  $\alpha$  whilst again varying  $\rho$  and  $d$ .

Figure 2 shows the size of the peak of infection and the final susceptible population as a function of the parameters  $\rho$  and  $d$  for different  $\alpha_T$  values. A common effect of behavioural feedback in epidemics is the reduction in the number of infected and increase in the final susceptible population [7, 12, 13]. We obtain a similar result in Figure 2(a) and 2(c). Increasing the information-mediated protection in the population,  $\rho$ , monotonically decreases the size of the peak of infection and hampers the spread of the disease, resulting in larger susceptible populations at the end of the epidemic.

In contrast, in Figures 2(b) and 2(d) we observe that larger densities of distrusting individuals monotonically increase the size of the peak of infection and lead to more overall infections, resulting in smaller



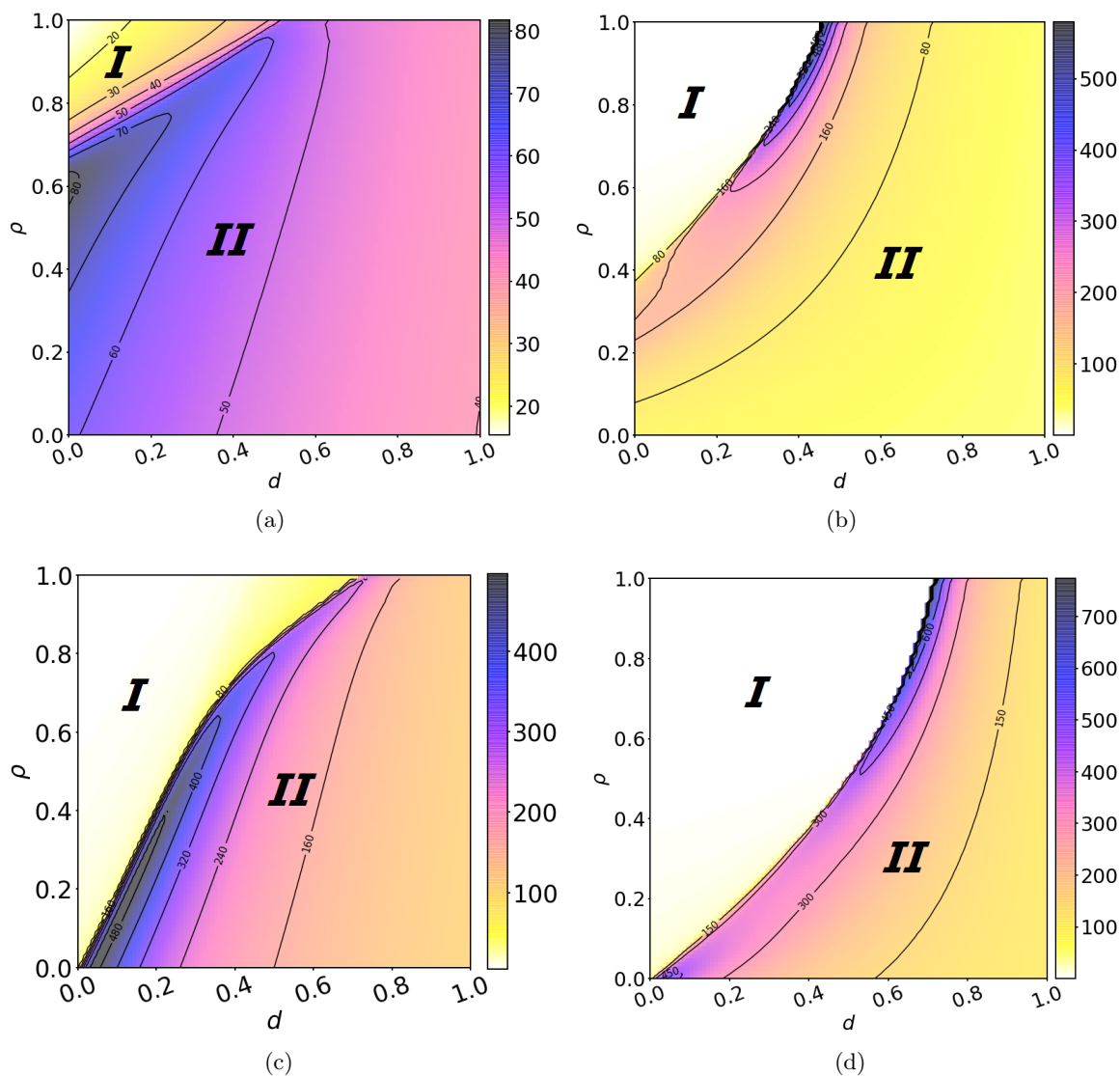
**Figure 2.** Faster information spread (increasing  $\alpha$ ) heightens model sensitivity to the strength of interventions ( $\rho$ ) and composition of the population ( $d$ ). Curves show effects of varying  $\rho$  and  $d$  on the peak infection size and the final susceptible population. (a) Total infections peak for varying  $\rho$  and fixed  $d = 0.3$ . (b) Total infections peak for fixed  $\rho = 0.8$  and varying  $d$ . (c) Final susceptible population for varying  $\rho$  and fixed  $d = 0.3$ . (d) Final susceptible population for fixed  $\rho = 0.8$  and varying  $d$ . Remaining parameters:  $\alpha_D = 1$ ,  $\lambda = 0.2$ ,  $\omega = 0.333$ ,  $\beta = 0.667$  and  $\sigma = 0.133$ .

susceptible populations at the end of the epidemic. This stems from the reluctance that this portion of the population has to abide by protective measures.

Our results highlight the importance of faster information spread in the trusting population, the use of non-pharmaceutical control measures, and the development of information campaigns to better inform the population and increase compliance with protective measures as a means to reduce disease spread and lighten the disease-associated healthcare burden.

Other important disease outbreak metrics are the expected duration of an epidemic and the time when the peak of infection occurs. Figure 1 already gave us an idea of how these quantities vary with  $\rho$  and  $d$ . To get a better picture of their effect, we solved the system of equations numerically again and measured the time until the peak of infection for different combinations of  $\rho$  and  $d$ , and different  $\alpha_T$  values. The results can be seen in Figure 3. The initial condition consisted of an almost entirely susceptible, unaware population with initial infected population density  $\frac{I_{T,\infty}}{N} = \frac{I_{D,\infty}}{N} = 5 \cdot 10^{-6}$ .

The first thing to notice from Figure 3 is the existence of the two regions mentioned earlier and the phase-transition, given by a curve  $\rho_c(d)$ . The first region (*I*), where  $\rho > \rho_c(d)$ , is the suppression regime, characterised by very low peaks of infection and times until the peak occurs that decrease as  $\rho$  increases. The second (*II*), where  $\rho < \rho_c(d)$ , is the mitigation regime, characterised by larger peaks of infection. In the mitigation regime, although the infection peak decreases as  $\rho$  increases, the times until the peak of infection increase, in contrast to the suppression regime. At the boundary, we observe a phase-transition, where the response to new infections is just strong enough to contain the increase



**Figure 3.** Timing of the infectious peak (colours) reveals the boundary between epidemic suppression and large outbreaks. First row:  $R_0 = 3.5$  ( $\beta = 0.467$ ), (a)  $\alpha_T = 0.5$  and (b)  $\alpha_T = 500$ . Second row:  $R_0 = 2$  ( $\beta = 0.267$ ), (c)  $\alpha_T = 0.5$  and (d)  $\alpha_T = 500$ . The remaining parameters are:  $\alpha_D = 1$ ,  $\lambda = 0.2$ ,  $\omega = 0.333$  and  $\sigma = 0.133$ .

in cases. At this boundary, we observe almost constant infection levels and, consequently, extended epidemics. The transition between the regions becomes more abrupt for larger  $\alpha_T$  values, i.e., when awareness transmission is much faster in the trusting population.

In addition to the two regions, we note the existence of a critical value  $d = d_M$  above which suppression of the disease is not possible. This significant result highlights the need for effective measures to combat the spread of misinformation and the importance of maintaining trust in order to contain and suppress the disease. Furthermore, increasing the speed of information spread in the trusting population reduces  $\rho_c(d)$ , increasing the area of parameter space for which suppression is achievable, in contrast with the increase of  $\alpha_D$ , which increases  $\rho_c(d)$ , reducing the suppression region of the parameter space.

In essence, we observe a second threshold in the model. The first one corresponds to the outbreak threshold, where the disease can initially spread in the unaware population, given by  $R_0 > 1$ . The second corresponds to the suppression threshold, where the protection elicited in the trusting population by the initial infection can reduce  $R_e$  to less than 1, hampering the spread of the disease. A second threshold due to preventive behaviour has been seen in the context of a SIS-like epidemiological model with awareness feedback on regular random networks and absence of fading mechanism [25], leading to a slow die-out of the epidemic and to a final population with a significant number of aware individuals. It has also been observed in [15, 16] in the case of a SIRS model with two states of aware and unaware individuals. In [15, 16], the authors observe that awareness changes the invasion conditions between a disease-free and endemic equilibrium, and can make it impossible for a disease to establish itself in the

population. In both cases, the endemic equilibria are characteristic of the chosen model, and the threshold can only be observed if awareness does not deteriorate through transmission or fading.

Naturally, the second threshold depends on the characteristics of the information spread dynamics,  $\alpha_T$  and  $\alpha_D$ , the link between the effectiveness of behavioural change  $\rho$ , and the population composition,  $d$ . This threshold can be reflected in the effectiveness of protection elicited by behaviour change  $\rho_c$  given the population composition (density of distrusting individuals,  $d$ ), and a fixed speed of information spread  $\alpha_T$  and  $\alpha_D$ . Smaller  $\rho_c$  values mean that less protective measures must be taken to reach the suppression threshold.

In the next section, we derive a theoretical expression for the value of  $\rho_c$  as function of  $d$  in the limit of fast information spread, i.e., large  $\alpha_T$  and  $\alpha_D$ . In this case, we observe a separation of timescales between disease and information spread, allowing an analytical treatment of the simplified model.

### 3.2. Estimates of the critical $\rho$ and density values under fast spread of information.

Although it is difficult to express the suppression threshold analytically in most cases, as it depends on the distribution of awareness in the population and its complicated dynamics, we can make use of extra assumptions, such as fast information spread and large population size  $N$ , to simplify the model. The procedure is similar to obtaining the basic reproduction number in most epidemiological models [26].

Consider an initially unaware population of size  $N$  with a small density of infected individuals. Additionally, consider the limit in which the spread of information happens almost instantaneously, i.e.  $\alpha_T \rightarrow \infty$  and  $\alpha_D \rightarrow \infty$ . This would be the case, for instance, where new cases are informed globally through the internet or mass media as soon as they are discovered. In this limit, there is a separation of time-scales between the information and disease spread dynamics. A brief moment after the beginning of the epidemic, the first infected individual refreshes their information, moving to index 0. Immediately, all remaining trusting individuals in the higher index levels move to index 1, while all distrusting individuals remain unaware. Note that once infected distrusting individuals refresh information, they immediately become unaware again. We can then assume that the distrusting population occupies only the  $S_{D,\infty}$ ,  $I_{D,\infty}$  and  $R_{D,\infty}$  subpopulations. In this scenario, the model simplifies to

$$\frac{dS_{T,1}}{dt} = -\frac{S_{T,1}}{N}(1-\rho)[(1-\rho)I_{T,1} + I_{D,\infty}]\beta, \quad (13)$$

$$\frac{dS_{D,\infty}}{dt} = -\frac{S_{D,\infty}}{N}[(1-\rho)I_{T,1} + I_{D,\infty}]\beta, \quad (14)$$

$$\frac{dI_{T,0}}{dt} = -\sigma I_{T,0} + \omega I_{T,1} - \lambda I_{T,0}, \quad (15)$$

$$\frac{dI_{T,1}}{dt} = -\sigma I_{T,1} - \omega I_{T,1} + \lambda I_{T,0} + \frac{S_{T,1}}{N}(1-\rho)[(1-\rho)I_{T,1} + I_{D,\infty}]\beta, \quad (16)$$

$$\frac{dI_{D,\infty}}{dt} = -\sigma I_{D,\infty} + \frac{S_{D,\infty}}{N}[(1-\rho)I_{T,1} + I_{D,\infty}]\beta, \quad (17)$$

where we have omitted the equations for the recovered populations. We drop the indices  $T$  and  $D$  for convenience as there is no shared information index between the two groups. As observed in the result section, we assume that the level of infected individuals is approximately constant (metastable) at the phase transition

$$\frac{dI}{dt} = \frac{dI_0}{dt} + \frac{dI_1}{dt} + \frac{dI_\infty}{dt} = 0. \quad (18)$$

We also assume metastability of all three infected groups. Setting  $\frac{dI_0}{dt} = 0$  gives  $I_0 = \left(\frac{\omega}{\sigma+\lambda}\right) I_1$ . Setting Eq. (17) to zero, and assuming a large population such that, after the initial phase of unrestricted transmission,  $\frac{S_\infty}{N} \approx d$ , where  $d$  is the initial density of distrusting, and  $\frac{S_1}{N} \approx 1-d$ , results in  $I_\infty = \frac{d(1-\rho)}{\sigma-\beta d} \beta I_1$ . Now, from Eq. (18),

$$0 = -\frac{\sigma}{\beta} I_1 \left( \frac{\omega}{\sigma+\lambda} + 1 + \frac{d(1-\rho)\beta}{\sigma-\beta d} \right) + (1-d)(1-\rho)^2 \left[ 1 + \frac{d\beta}{\sigma-\beta d} \right] I_1 + I_1 d(1-\rho) \left[ 1 + \frac{d\beta}{\sigma-\beta d} \right].$$

Denoting  $x = (1-\rho)$  and  $y = \frac{\beta}{\sigma-\beta d}$ ,

$$x^2(1-d)(1+dy) + xd \left( 1 + dy - \frac{y}{R_0} \right) - \frac{1}{R_0} \left( 1 + \frac{\omega}{\sigma+\lambda} \right) = 0,$$

$$x^\pm(d) = \frac{-d \left( 1 + dy - \frac{y}{R_0} \right) \pm \sqrt{d^2 \left( 1 + dy - \frac{y}{R_0} \right)^2 + \frac{4}{R_0} \left( 1 + \frac{\omega}{\sigma+\lambda} \right) (1-d)(1+dy)}}{2(1-d)(1+dy)}. \quad (19)$$



We then obtain the critical protection level  $\rho_c$  as a function of the distrusting population density  $d$ ,  $\rho_c(d) = 1 - x(d)$ . For  $\rho < \rho_c(d)$  we observe mitigation of the disease spread, whilst for  $\rho > \rho_c(d)$  the disease is suppressed.

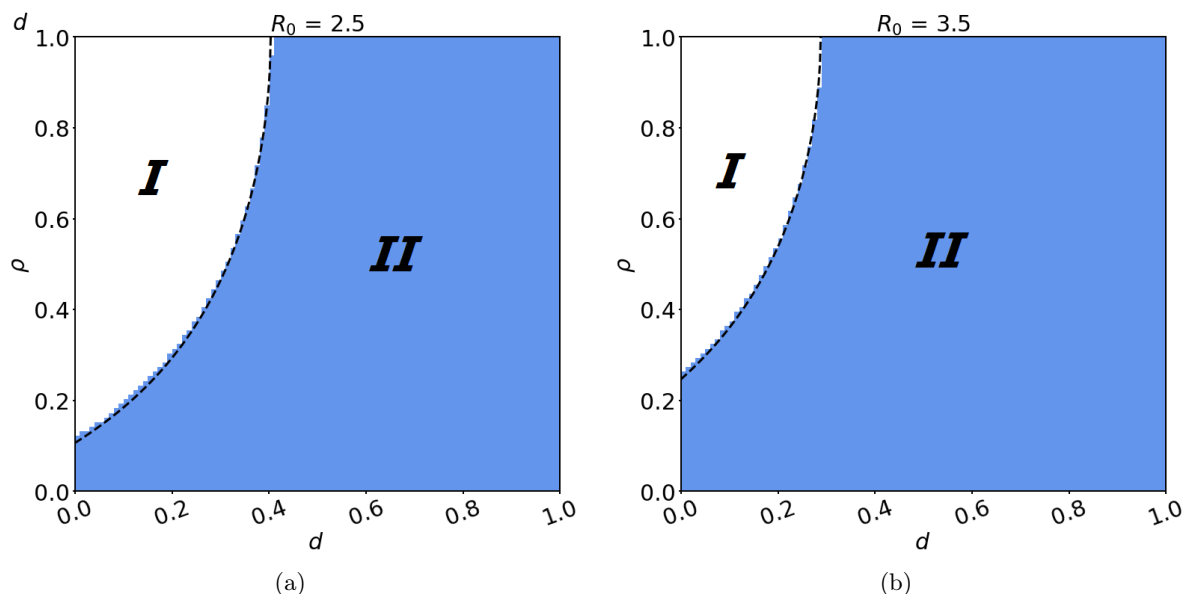
From the expression above, we note that there is a critical distrusting density  $d_M = \frac{\sigma}{\beta} = \frac{1}{R_0}$  above which suppression of the disease is not feasible irrespective of the level of information-mediated protection,  $\rho$ . Solutions for  $d > d_M$  yield negative  $I_\infty$  population sizes and are, therefore, not physical. Similarly, we are interested in solutions for which  $0 < \rho < 1$ , or equivalently,  $0 < x < 1$ . Notice that, for  $d < d_M$ , the  $x^+$  solution is always greater than zero, giving  $\rho_c < 1$ . Hence, there always exists a region for  $d < d_M$  in which suppression of the disease is achievable for sufficiently large  $\rho$ . Notice that, if  $R_0 < 1$ , then  $d_M > 1$ . Hence, suppression is always achieved in this case. This is consistent with the threshold for an initial outbreak  $R_0 > 1$ .

In particular, for  $d = 0$ ,

$$\rho_c(0) = 1 - \sqrt{\frac{\sigma}{\beta} \left( 1 + \frac{\omega}{\sigma + \lambda} \right)}, \quad (20)$$

gives us conditions where  $\rho_c(0) > 0$  and we can see a phase transition from mitigation to suppression for increasing  $\rho$  in the absence of distrusting individuals. If  $\rho_c(0) < 0$ , then suppression is achieved for all  $\rho$  values, and we do not observe the transition as  $\rho$  increases. From Eq. (19), and more easily from Eq. (20), we can see that increasing  $\omega$  while keeping the remaining parameters constant reduces  $\rho_c(d)$ , since more information is being produced and transmitted, which increases awareness in the population. In contrast, increasing  $\lambda$  increases  $\rho_c(d)$  as the information is forgotten faster. Similarly, increasing  $\sigma$  while keeping  $R_0$  constant increases  $\rho_c(d)$ , as the probability that individuals recover before realising they were infected increases.

Figure 4 shows the regions of  $d - \rho$  parameter space corresponding to suppression (*I*) and mitigation (*II*) obtained through the numerical solution of the complete model (with large  $\alpha$  and  $I_{T,\infty}/N = I_{D,\infty}/N = 5 \cdot 10^{-6}$ ), and the theoretically predicted boundary (dashed black line), Eq. (19), for two values of  $\beta$  (and corresponding two values of  $R_0$ ). Good agreement was also observed for different combinations of  $\beta$  and  $\sigma$  (not shown), although discrepancies arise as the rate of information spread is reduced. Nonetheless, Eq.(19) shows the existence of regions of suppression and mitigation and a critical density of distrusting individuals above which suppression is not achievable, emphasising the importance of effective measures to combat misinformation and increase the population's compliance with disease-control measures.



**Figure 4.** The agreement between the theoretical prediction (dashed lines) of the observed phase transition and the suppression/mitigation regions obtained through the numerical solutions of Eqs.(3-10) for  $\alpha_T = \alpha_D = 10^6$ ,  $\lambda = 0.2$ ,  $\omega = 0.33$ ,  $\sigma = 0.133$ , and (a)  $\beta = 0.333$  ( $R_0 = 2.5$ ) and (b)  $\beta = 0.467$  ( $R_0 = 3.5$ ). The white area (*I*) correspond to the region where disease spread is suppressed. The area in blue (*II*) correspond to the region where disease spread is only mitigated by behavioural feedback.

#### 4. DISCUSSION

Mathematical models are pivotal to policy-making and managing epidemics, giving insight into important aspects of the dynamics of a disease that can facilitate its control and predict possible outcomes. In the past, epidemiological models failed to account for the important aspect of behavioural responses to diseases in human populations. More recently, prominent approaches focusing on the spread of awareness and its interplay with disease dynamics have been developed [5].

In this article, we developed a model that merged epidemiological dynamics with information spread and behavioural change interactions of both trusting and distrusting individuals in a population. This model is an extension of previous work in this area which considered only one class of individuals [13]. Using our model, we were able to highlight the mechanisms by which the burden of epidemics can be lightened.

Our results show that the surge in awareness, arising from fast spread of information in the trusting population, can reduce the peak of infection and lead to a larger population that remains uninfected by the end of the outbreak. A similar effect is observed by increasing the effectiveness of protective measures taken by well-informed individuals. We also observed a phase transition between mitigation and suppression regimes. In the mitigation regime, the behavioural reaction and reduction in transmissibility arising in the population is not sufficient to reduce the effective reproduction number below one. Hence, the observed effect in the infection curve is a lower and delayed peak. However, if the response to new cases is sufficiently strong, a fine balance between new infections and a reduction in transmissibility can be achieved. In this case, we observe an almost constant level of infection over time, leading to extended epidemics. We then observe a phase transition, in which further increasing the effectiveness of the measures to control the spread of the disease leads to a regime of suppression. In contrast, larger densities of distrusting individuals in the population limit the effectiveness of the reduced transmissibility in the trusting population, resulting in the existence of a critical density  $d_M$  from which the suppression regime is not achievable.

Although a complete and simple analytical characterisation of the suppression threshold is difficult in general, as it depends on the complex dynamics of the system, we were able to write an expression for  $\rho_c(d)$  in the limit of fast information transmission and assuming a large population size. The expression agreed with the numerically observed boundary in the corresponding parameter regime, and allowed us to obtain the maximum proportion of distrusting individuals for which suppression of the outbreak is achievable. A possibly fruitful direction for further research to obtain approximations for realistic information transmission parameter values, but that still exhibit slow and fast dynamics, is the analysis of the stochastic version of the system using the framework developed in [27] that allows for dimension reduction in stochastic dynamical systems with a separation of timescales. The method is exact in the limit of well-separated fast-and-slow dynamics and small noise, and found to be a reasonable approximation over a sensible parameter range.

The deterministic model developed has the underlying assumption of a well-mixed population, for which the explicit spatial structure of the system is disregarded. Previous work on the effects of spatial structure in the spread of infectious diseases has shown its importance in the invasion threshold for epidemics [28], and the development of control policies [29–32]. The analysis of epidemiological models with awareness feedback on networks with homogeneous populations has been an active area of research [10, 33–38]. However, these models disregard the heterogeneous behaviour presented here. Thus, the inclusion of spatial dynamics in our model is another prospective direction for further work.

#### ACKNOWLEDGEMENTS

A.S. is supported by a scholarship from the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), under the project EP/S022945/1. T.R. acknowledges the support of the Royal Society RGF\EA\180242. This research made use of the Balena High Performance Computing (HPC) Service at the University of Bath.

#### REFERENCES

- [1] J. N. Hays, *Epidemics and pandemics: Their impacts on human history*, ABC-CLIO, **2006**.
- [2] J. Howard, A. Huang, Z. Li, Z. Tufekci, V. Zdimal, H.-M. van der Westhuizen, A. von Delft, A. Price, L. Fridman, L.-H. Tang, V. Tang, G. L. Watson, C. E. Bax, R. Shaikh, F. Questier, D. Hernandez, L. F. Chu, C. M. Ramirez, A. W. Rimoin, “An evidence review of face masks against COVID-19”, *Proceedings of the National Academy of Sciences* **2021**, *118*, e2014564118.

- [3] D. K. Chu, E. A. Akl, S. Duda, K. Solo, S. Yaacoub, H. J. Schünemann, D. K. Chu, E. A. Akl, A. El-harakeh, A. Bognanni, T. Lotfi, M. Loeb, A. Hajizadeh, A. Bak, A. Izovich, C. A. Cuello-Garcia, C. Chen, D. J. Harris, E. Borowiack, F. Chamseddine, F. Schünemann, G. P. Morgano, G. E. U. M. Schünemann, G. Chen, H. Zhao, I. Neumann, J. Chan, J. Khabisa, L. Hneiny, L. Harrison, M. Smith, N. Rizk, P. G. Rossi, P. AbiHanna, R. El-khoury, R. Stalteri, T. Baldeh, T. Piggott, Y. Zhang, Z. Saad, A. Khamis, M. Reinap, S. Duda, K. Solo, S. Yaacoub, H. J. Schünemann, “Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis”, *The Lancet* **2020**, *395*, 1973–1987.
- [4] E. Mahase, “Covid-19: Was the decision to delay the UK’s lockdown over fears of “behavioural fatigue” based on evidence?”, *BMJ* **2020**, m3166.
- [5] S. Funk, M. Salathé, V. A. A. Jansen, “Modelling the influence of human behaviour on the spread of infectious diseases: a review”, *Journal of The Royal Society Interface* **2010**, *7*, 1247–1256.
- [6] T. C. Reluga, “Game Theory of Social Distancing in Response to an Epidemic”, *PLoS Computational Biology* **2010**, *6*, (Ed.: C. T. Bergstrom), e1000793.
- [7] R. C. Tyson, S. D. Hamilton, A. S. Lo, B. O. Baumgaertner, S. M. Krone, “The Timing and Nature of Behavioural Responses Affect the Course of an Epidemic”, *Bulletin of Mathematical Biology* **2020**, *82*.
- [8] J. M. Epstein, J. Parker, D. Cummings, R. A. Hammond, “Coupled Contagion Dynamics of Fear and Disease: Mathematical and Computational Explorations”, *PLoS ONE* **2008**, *3*, (Ed.: A. P. Galvani), e3955.
- [9] J. M. Epstein, E. Hatna, J. Crodelle, Coupled Contagion: A Two-Fears Epidemic Model, **2021**.
- [10] A. Rizzo, M. Frasca, M. Porfiri, “Effect of individual behavior on epidemic spreading in activity-driven networks”, *Phys. Rev. E* **2014**, *90*, 042801.
- [11] K. A. Kabir, K. Kuga, J. Tanimoto, “Analysis of SIR epidemic model with information spreading of awareness”, *Chaos Solitons & Fractals* **2019**, *119*, 118–125.
- [12] J. S. Weitz, S. W. Park, C. Eksin, J. Dushoff, “Awareness-driven behavior changes can shift the shape of epidemics away from peaks and toward plateaus, shoulders, and oscillations”, *Proceedings of the National Academy of Sciences* **2020**, *117*, 32764–32771.
- [13] S. Funk, E. Gilad, C. Watkins, V. A. A. Jansen, “The spread of awareness and its impact on epidemic outbreaks”, *Proceedings of the National Academy of Sciences* **2009**, *106*, 6872–6877.
- [14] L. J. Allen, “A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis”, *Infectious Disease Modelling* **2017**, *2*, 128–142.
- [15] S. Funk, E. Gilad, V. Jansen, “Endemic disease, awareness, and local behavioural response”, *Journal of Theoretical Biology* **2010**, *264*, 501–509.
- [16] I. Z. Kiss, J. Cassell, M. Recker, P. L. Simon, “The impact of information transmission on epidemic outbreaks”, *Mathematical Biosciences* **2010**, *225*, 1–10.
- [17] M. J. Hornsey, E. A. Harris, K. S. Fielding, “The psychological roots of anti-vaccination attitudes: A 24-nation investigation.”, *Health Psychology* **2018**, *37*, 307–315.
- [18] K. C. London, The three groups reacting to life under lockdown, **2020**.
- [19] K. C. London, Trust in LOCKDOWN sceptics linked to arguments about Covid vaccine, **2021**.
- [20] J. Murphy, F. Vallières, R. P. Bentall, M. Shevlin, O. McBride, T. K. Hartman, R. McKay, K. Bennett, L. Mason, J. Gibson-Miller, L. Levita, A. P. Martinez, T. V. A. Stocks, T. Karatzias, P. Hyland, “Psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in Ireland and the United Kingdom”, *Nature Communications* **2021**, *12*.
- [21] J. Milošević Đorđević, S. Mari, M. Vdović, A. Milošević, “Links between conspiracy beliefs, vaccine knowledge, and trust: Anti-vaccine behavior of Serbian adults”, *Social Science & Medicine* **2021**, *277*, 113930.
- [22] R. M. Anderson, R. M. May, *Infectious diseases of humans. dynamics and ontrol*, Oxford University Press, **1991**.
- [23] O. Diekmann, H. J. A. P., *Mathematical epidemiology of infectious diseases: Model building, anaysis and interpretation*, John Wiley & Sons, **2000**.
- [24] A. Catching, S. Capponi, M. T. Yeh, S. Bianco, R. Andino, “Examining the interplay between face mask usage, asymptomatic transmission, and social distancing on the spread of COVID-19”, *Scientific Reports* **2021**, *11*.
- [25] D. Juher, I. Z. Kiss, J. Saldaña, “Analysis of an epidemic model with awareness decay on regular random networks”, *Journal of Theoretical Biology* **2015**, *365*, 457–468.
- [26] P. van den Driessche, J. Watmough, “Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission”, *Mathematical Biosciences* **2002**, *180*, 29–48.
- [27] T. L. Parsons, T. Rogers, “Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: a constructive approach with examples from theoretical biology”, *Journal of Physics A: Mathematical and Theoretical* **2017**, *50*, 415601.

- [28] M. J. Keeling, “The effects of local spatial structure on epidemiological invasions”, *Proceedings of the Royal Society of London. Series B: Biological Sciences* **1999**, *266*, 859–867.
- [29] M. J. Keeling, M. E. J. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-Topping, D. T. Haydon, S. J. Cornell, J. Kappey, J. Wilesmith, B. T. Grenfell, “Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape”, *Science* **2001**, *294*, 813–817.
- [30] N. M. Ferguson, C. A. Donnelly, R. M. Anderson, “Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain”, *Nature* **2001**, *413*, 542–548.
- [31] N. M. Ferguson, D. A. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meechai, S. Iamsirithaworn, D. S. Burke, “Strategies for containing an emerging influenza pandemic in Southeast Asia”, *Nature* **2005**, *437*, 209–214.
- [32] I. Gudelj, K. White, “Spatial heterogeneity, social structure and disease dynamics of animal populations”, *Theoretical Population Biology* **2004**, *66*, 139–149.
- [33] S. Bansal, B. T. Grenfell, L. A. Meyers, “When individual behaviour matters: homogeneous and network models in epidemiology”, *Journal of The Royal Society Interface* **2007**, *4*, 879–891.
- [34] F. D. Sahneh, C. M. Scoglio in 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), IEEE, **2012**.
- [35] C. Granell, S. Gómez, A. Arenas, “Dynamical Interplay between Awareness and Epidemic Spreading in Multiplex Networks”, *Phys. Rev. Lett.* **2013**, *111*, 128701.
- [36] F. Sahneh, F. Chowdhury, G. Brase, C. Scoglio, “Individual-based Information Dissemination in Multilayer Epidemic Modeling”, *Mathematical Modelling of Natural Phenomena* **2014**, *9*, (Eds.: J. Hyman, F. Milner, J. Saldaña), 136–152.
- [37] J.-Q. Kan, H.-F. Zhang, “Effects of awareness diffusion and self-initiated awareness behavior on epidemic spreading - An approach based on multiplex networks”, *Communications in Nonlinear Science and Numerical Simulation* **2017**, *44*, 193–203.
- [38] P. Hu, L. Ding, X. An, “Epidemic spreading with awareness diffusion on activity-driven networks”, *Phys. Rev. E* **2018**, *98*, 062322.