

# A generalized model of social and biological contagion

P.S. Dodds<sup>a,\*</sup>, D.J. Watts<sup>b,c</sup>

<sup>a</sup>*Institute for Social and Economic Research and Policy, Columbia University, 420 West 118th Street, MC 3355, New York, NY 10027, USA*

<sup>b</sup>*Department of Sociology, Columbia University, 1180 Amsterdam Avenue, New York, NY 10027, USA*

<sup>c</sup>*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*

Received 27 May 2004; received in revised form 15 September 2004; accepted 15 September 2004

Available online 6 November 2004

## Abstract

We present a model of contagion that unifies and generalizes existing models of the spread of social influences and microorganismal infections. Our model incorporates individual memory of exposure to a contagious entity (e.g. a rumor or disease), variable magnitudes of exposure (dose sizes), and heterogeneity in the susceptibility of individuals. Through analysis and simulation, we examine in detail the case where individuals may recover from an infection and then immediately become susceptible again (analogous to the so-called SIS model). We identify three basic classes of contagion models which we call *epidemic threshold*, *vanishing critical mass*, and *critical mass* classes, where each class of models corresponds to different strategies for prevention or facilitation. We find that the conditions for a particular contagion model to belong to one of these three classes depend only on memory length and the probabilities of being infected by one and two exposures, respectively. These parameters are in principle measurable for real contagious influences or entities, thus yielding empirical implications for our model. We also study the case where individuals attain permanent immunity once recovered, finding that epidemics inevitably die out but may be surprisingly persistent when individuals possess memory.

© 2004 Elsevier Ltd. All rights reserved.

**Keywords:** Contagion; Epidemic; Memory; Threshold; Dose response; Critical mass; Universality; Bifurcation

## 1. Introduction

Contagion, in its most general sense, is the spreading of an entity or influence between individuals in a population, via direct or indirect contact. Contagion processes therefore arise broadly in the social and biological sciences, manifested as, for example the spread of infectious diseases (Murray, 2002; Daley and Gani, 1999; Anderson and May, 1991; Brauer and Castillo-Chávez, 2001; Diekmann and Heesterbeek, 2000; Hethcote, 2000) and computer viruses, the diffusion of innovations (Coleman et al., 1966; Valente, 1995; Rogers, 1995), political upheavals (Lohmann, 1994), and the dissemination of religious doctrine (Stark, 1996; Montgomery, 1996). Existing mathematical models of contagion, while motivated in a variety

of ways depending on the application at hand, fall into one of only two broad categories, where the critical distinction between these categories can be explained in terms of the interdependencies between successive contacts; that is, the extent to which the effect of an exposure to a contagious agent is determined by the presence or absence of previous exposures.

The standard assumption in all mathematical models of infectious disease spreading (for example, the classic SIR model, Kermack and McKendrick, 1927; Murray, 2002), and also in some models of social contagion (Goffman and Newill, 1964; Daley and Kendall, 1965; Bass, 1969), is that there is no interdependency between contacts; rather, the infection probability is assumed to be independent and identical across successive contacts. All such models fall into a category that we call *independent interaction models*. By contrast, what we call *threshold models* assert that an individual can only become infected when a certain critical number of

\*Corresponding author. Tel.: +1 212 854 9647.

E-mail address: [peter.dodds@columbia.edu](mailto:peter.dodds@columbia.edu) (P.S. Dodds).

exposures has been exceeded, at which point infection becomes highly probable. The presence of a threshold corresponds to interdependencies of an especially strong nature: contacts that occur when an individual is near its threshold are extremely consequential while others have little or no effect. Threshold models are often used to describe social contagion (e.g. the spreading of fads or rumors), where individuals either deterministically (Schelling, 1973; Granovetter, 1978; Watts, 2002) or stochastically (Bikhchandani et al., 1992; Banerjee, 1992; Morris, 2000; Brock and Durlauf, 2001) “decide” whether or not to adopt a certain behavior based in part or in whole on the previous decisions of others.

An alternative way to think about the interdependence of successive events is in terms of memory: threshold models implicitly assume the presence of memory while independent interaction models assume (again implicitly) that the infection process is memoryless.<sup>1</sup> Neither class of model, however, is able to capture the dynamics of contagion processes that possess an intermediate level of interdependency, or equivalently a variable emphasis on memory. Furthermore, the relationship between interdependent interaction models, threshold models, and any possible intermediate models is unclear. Motivated by these observations, our model seeks to connect threshold and independent interaction models both conceptually and analytically, and explore the classes of contagion models that lie between them. Such an analysis is clearly relevant to the problems of social contagion, in which memory of past events obviously plays some role, but one that may be less strong than is assumed by most threshold models. However, it may also be relevant to biological disease spreading models, which, to our knowledge, have previously not questioned the assumption of independence between successive exposures. While the independence assumption is indeed plausible, it has not been demonstrated empirically, and little enough is understood of the dynamics of immune system response that the alternative—persistent sub-critical doses of an infectious agent combining to generate a critical dose—cannot be ruled out. Furthermore, memory effects are known to be inherent to certain kinds of immune system responses, such as allergic response. Hence if, as we indeed show, only a slight departure from complete independence is required to alter the corresponding collective dynamics, then our approach may also shed light on the spread of infectious diseases.

<sup>1</sup>Note that independent interaction models typically incorporate a state of immunity, which reflects a type of memory, but not the kind with which we are chiefly concerned here—namely, the memory of a contagious influence or entity prior to an infection occurring.

## 2. Description of model

Our model, aspects of which we have reported in brief elsewhere (Dodds and Watts, 2004), comprises a population of  $N$  individuals, each of which is assumed to occupy one of three states: *susceptible* (S); *infected* (I); or *removed* (R). Fig. 1 provides a schematic representation of our model which we describe as follows. At each (discrete) time step  $t$ , each individual  $i$  comes into contact with another individual chosen uniformly at random from the population. If the contact is infected—an event that occurs with probability  $\phi_t$ , the fraction of infectives in the population—then with probability  $p$ ,  $i$  receives a ‘dose’  $d$  drawn from a fixed dose-size distribution  $f$ ; else  $i$  receives a dose of size 0. We call a successful transmission of a positive dose an *exposure* and  $p$  the *exposure probability*. Individuals carry a *memory* of doses received from their last  $T$  contacts and we denote the sum of individual  $i$ ’s last  $T$  doses ( $i$ ’s *dose count*) at the  $t$ th time step by

$$D_{t,i} = \sum_{t'=t-T+1}^t d_{t',i}. \tag{1}$$

If  $i$  is in the susceptible state, then it becomes infected once  $D_{t,i}$  exceeds  $i$ ’s *dose threshold*  $d_i^*$ , where  $d_i^*$  is drawn from a given distribution  $g$  (dose thresholds do not change with time). Note that we differentiate exposure from infection, the latter being the possible result of one or more exposures and only occurring once a susceptible individual’s threshold has been equaled or exceeded. Having become infected, an individual remains in state I until its dose count drops below its threshold, at which point it recovers with probability  $r$  at each time step. Once recovered, an individual returns to being susceptible with probability  $\rho$ , again at each time step.

The probability  $P_{inf}$  that a susceptible individual who comes into contact with  $K$  infected individuals in  $T$  time steps will become infected is therefore given by

$$P_{inf}(K) = \sum_{k=1}^K \binom{K}{k} p^k (1-p)^{K-k} P_k, \tag{2}$$

where  $K = 1, \dots, T$ , and

$$P_k = \int_0^\infty dd^* g(d^*) \int_{d=d^*}^\infty dd f^{k*}(d). \tag{3}$$

Both  $P_{inf}(K)$  and  $P_k$  are important quantities in our model. The quantity  $P_k$  is the expected fraction of a population that will be infected by  $k$  exposures, where the distribution  $f^{k*}(d)$ , the  $k$ -fold convolution of  $f$ , is the probability that the sum of  $k$  doses will be equal to  $d$ . The infection probability  $P_{inf}(K)$  gives, in effect, a *dose-response curve* (Haas, 2002) averaged over all

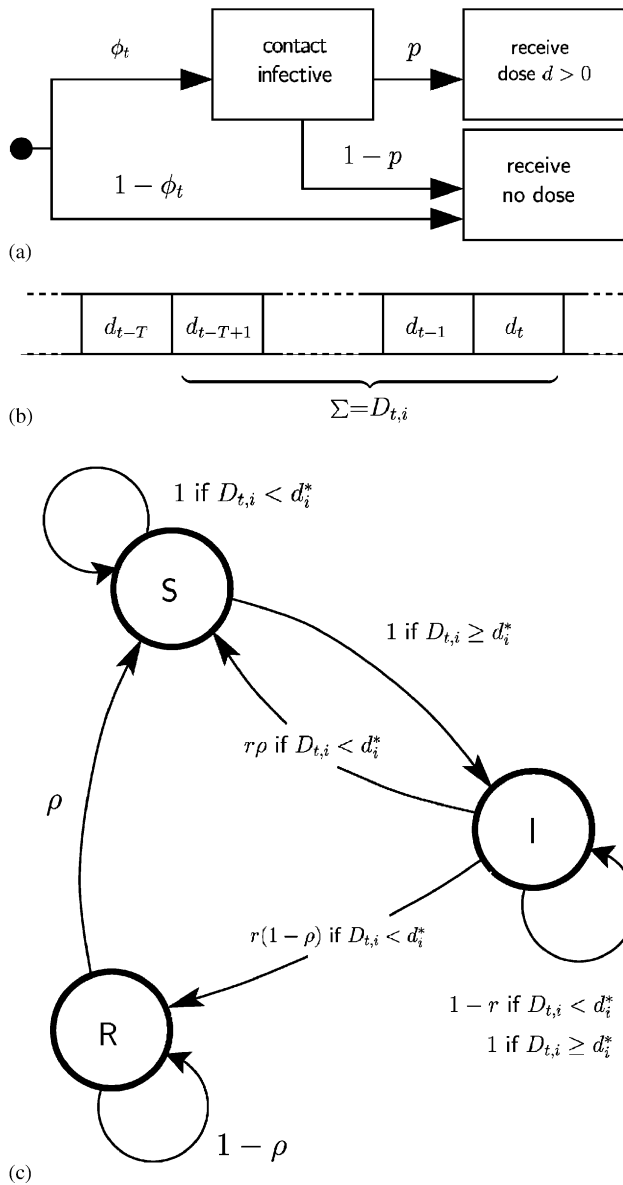


Fig. 1. (a) Representation of how an individual’s dose at time  $t$  is determined. At each time step, each individual, *regardless* of its state, contacts one other randomly chosen individual. The probability that individual  $i$  contacts an infective is then  $\phi_t$ , i.e. the current fraction of infectives in the population. If the contact is infected then with probability  $p$ ,  $i$  is exposed to the contagious entity and receives a dose  $d$  drawn from a fixed distribution  $f$ . Otherwise and also if the contacted individual is not infected in the first place (occurring with probability  $1 - \phi_t$ ),  $i$  receives a dose of zero size. (b) Individual  $i$  then updates its dose count  $D_{t,i}$  by ‘forgetting’ the dose it received  $T + 1$  time steps ago and incorporating the current dose [Eq. (1)]. (c) Transition probabilities for individual  $i$  cycling through the three states S (susceptible), I (infected), and R (recovered). If  $i$  is in the susceptible state, it becomes infected with probability 1 once its dose count  $D_{t,i}$  exceeds its threshold  $d_i^*$ , otherwise it remains susceptible. If  $i$  becomes infected, then whenever  $i$ ’s dose count drops below its threshold, it has a probability  $r$  of recovering in each time step. Once  $i$  is in state R (where it is immune to infection), it becomes susceptible again with probability  $\rho$  again in each time step. Note that if  $r = \rho = 1$ , infected individuals whose dose count falls below their threshold immediately return to the susceptible state.

members of the population and also the distribution of dose sizes (where we note that  $K$  contacts with infected individuals will result in  $k$  actual exposures with probability  $\binom{K}{k} p^k (1 - p)^{K-k}$ ). Fig. 2 shows examples of dose–response curves, calculated from Eq. (2) for four configurations of the model. The plots correspond to (a) independent interaction, (b) deterministic threshold, and, in both (c) and (d), stochastic threshold models. For the independent interaction example, a single exposure is needed to generate an infection and so exposures effectively act independently. The deterministic threshold example incorporates uniform dose sizes and thresholds, and when the probability of an exposure is set to  $p = 1$ , the response becomes deterministic (individuals are always infected when their dose count is met or exceeded and never otherwise); but now the threshold can only be exceeded by multiple infections. The two stochastic cases generalize the deterministic case by allowing (c) dose sizes to be heterogeneous, and (d) both dose sizes and thresholds to be heterogeneous.

We explore the behavior of our model with respect to three qualitative types of dynamics: (1) permanent removal ( $\rho = 0$ ) dynamics, analogous to so-called SIR models in mathematical epidemiology in which individuals either die or acquire permanent immunity; (2) temporary removal ( $1 > \rho > 0$ ) dynamics, analogous to SIRS models where recovered individuals become susceptible again after a certain period of immunity; and (3) instantaneous replacement ( $\rho = 1$ ) dynamics, analogous to SIS models, where infected individuals immediately become susceptible again upon recovery. Chicken pox, for example, would correspond to SIR-type contagion, while the common cold resembles the SIRS case. Because of its simplicity, we obtain the majority of our analytic results for the somewhat special SIS case ( $\rho = 1$ ). However, our main findings for the SIS case have analogs in the more complicated SIRS and SIR cases which we investigate with numerical simulations. Furthermore, while the assumption of instantaneous re-susceptibility is probably not appropriate in the case of contagious biological agents (where recovery is generally associated with some period of immunity), it may well be approximately true for contagious social influences, such as “social smoking,” where an individual, having quit, can restart immediately. A summary of the main parameters of the model and their definitions is provided in Table 1.

We structure the remainder of the paper as follows. In Section 3, through analysis and simulations, we examine in detail the SIS ( $\rho = 1$ ) version of the model for a homogeneous population, where by “homogeneous,” we mean that all doses are equal and of unit size [i.e.  $f(d) = \delta(d - 1)$ ], and that all individuals have the same threshold [i.e.  $g(d^*) = \delta(d^* - \bar{d}^*)$ ]. For homogeneous populations, we find that only two universal classes of

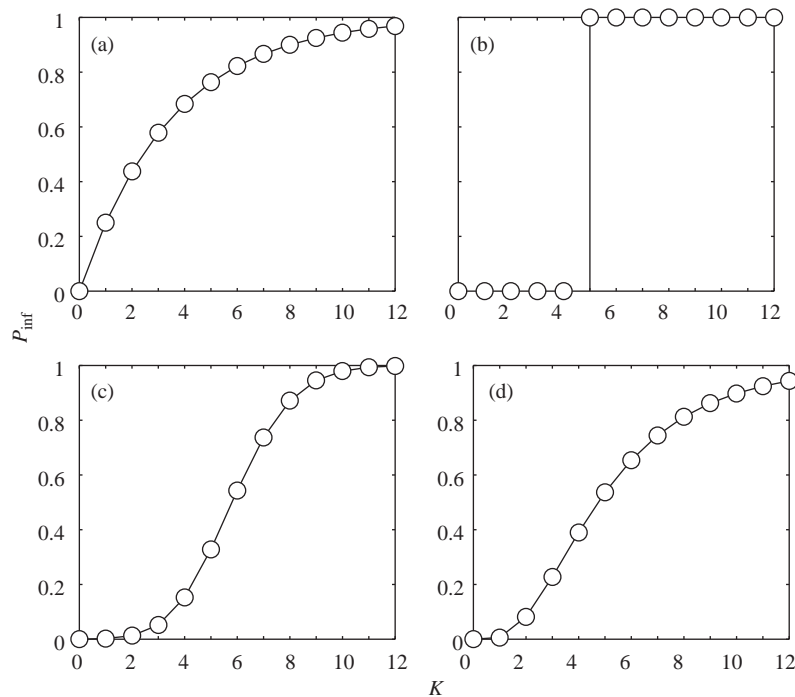


Fig. 2. Examples of dose–response, i.e.  $P_{inf}(K)$ , the probability an individual becomes infected due to  $K$  exposures in the last  $T(= 12)$  time steps [see Eq. (2)]. The plots correspond to (a) independent interaction (or disease-like) models with homogeneous dose sizes and thresholds [ $p = 0.25$ , dose distribution  $f(d) = \delta(d - 1)$ , and threshold distribution  $g(d^*) = \delta(d^* - 1)$ ]; (b) deterministic threshold models with homogeneous dose sizes and thresholds [ $p = 1$ ,  $f(d) = \delta(d - 1)$ , and  $g(d^*) = \delta(d^* - 5)$ ]; (c) stochastic threshold models with heterogeneous doses and homogeneous thresholds [ $p = 0.9$ , doses distributed lognormally with unit mean and standard deviation 0.5, and  $g(d^*) = \delta(d^* - 5)$ ]; and (d), same as (c) but with heterogeneity in thresholds introduced via a lognormal distribution with mean of 5 as per (c) and standard deviation of 10.

Table 1  
Summary of main model parameters and definitions

$T$	Length of memory window
$p$	Probability of exposure given contact with infective
$r$	Probability of moving from infected to recovered state
$\rho$	Probability of moving from immune to susceptible state
$f(d)$	Distribution of dose sizes $d$
$g(d^*)$	Distribution of individual thresholds $d^*$
$\bar{d}^*$	Uniform threshold of homogeneous population
$\phi_t$	Fraction of population infected at time $t$
$\phi^*$	Steady-state fraction of population infected

dynamics are possible: (1) *epidemic threshold dynamics*,<sup>2</sup> according to which initial outbreaks either die out or else infect a finite fraction of the population, depending on whether or not the infectiousness  $p$  exceeds a specific critical value  $p_c$ ; and (2) *critical mass dynamics* according to which a finite fraction of the population can only ever be infected in equilibrium if the initial outbreak size

itself constitutes a finite “critical mass.” Although homogeneity is a restrictive assumption for biological or social contagion, it provides a useful special case in that it illuminates the basic intuitions required to understand the more general, heterogeneous case. In Section 4, we relax the homogeneity assumption and move to the richer and more realistic case of arbitrarily distributed dose sizes and individual thresholds. Here, we find that three universal classes of contagion models are possible. In addition to the epidemic threshold and critical mass classes that carry over from the homogeneous case, we also find an intermediate class of *vanishing critical mass dynamics* in which the size of the required critical mass diminishes to zero for  $p < 1$ . Furthermore, we determine where the transitions between these classes occur and also the conditions required for more complicated kinds of contagion models to arise. Subsequently, in Section 5, we explore the SIRS and SIR versions of the model, finding behavior that in many ways resembles that of the simpler SIS case. In the SIR case, for example, it is necessarily true that all epidemics eventually burn themselves out (because for  $\rho = 0$ , the removed condition is an absorbing state). However, we find that the presence of memory may cause an epidemic to persist for a surprisingly long time. In Section 6, we conclude our analysis, discussing briefly the implications of our

<sup>2</sup>In this paper, we use the word ‘threshold’ in two terms: ‘threshold models’ and ‘epidemic threshold models.’ At the risk of some confusion, we have done so to maintain consistency with the two distinct literatures we are connecting, sociology and mathematical epidemiology. Threshold models of sociology refer to individual level thresholds whereas the term epidemic threshold refers to the critical reproduction number of a disease, above which an outbreak is assured.

findings for stimulating or retarding different contagion processes. Finally, in the appendices we provide detailed derivations of the analytical results from Sections 3 and 4.

### 3. Homogeneous SIS contagion models

#### 3.1. Epidemic threshold models

We begin our analysis with a simple non-trivial case of our model, for which we assume that individuals are: identical (i.e. the population is homogeneous); have no memory of past interactions ( $T = 1$ ); and, upon recovery from infection, immediately return to the susceptible state (i.e.  $\rho = 1$ ). In this limit our model coincides with the SIS version of the traditional Kermack–McKendrick model (Kermack and McKendrick, 1927; Murray, 2002), as individuals with no memory necessarily become infected upon exposure to a single infected individual [ $g(d^*) = \delta(d^* - 1)$ ]. For a specified recovery time ( $r \leq 1$ ) (which again is identical across all individuals), the fraction of infected individuals at time  $t$ ,  $\phi_t$ , evolves according to

$$\phi_{t+1} = p\phi_t + \phi_t(1 - p\phi_t)(1 - r). \quad (4)$$

The first term on the right is the fraction of individuals newly infected between time  $t$  and  $t + 1$ , regardless of whether or not they were infected beforehand (the model allows for individuals to recover and be reinfected within one time step). The second term is the fraction of individuals who were infected in the preceding time step, were not infected between time  $t$  and  $t + 1$ , and did not recover.

Setting  $\phi_t = \phi^*$  in Eq. (4), we find the stable fixed points of the model as a function of  $p$  are given by

$$\phi^* = \begin{cases} 0 & \text{for } 0 \leq p \leq r, \\ \frac{1 - r/p}{1 - r} & \text{for } r < p \leq 1. \end{cases} \quad (5)$$

Furthermore, a single unstable set of fixed points is found along  $\phi^* = 0$  for  $r < p \leq 1$ . Thus, the standard memoryless SIS model exhibits an epidemic threshold (Murray, 2002) at  $p = p_c$ , as displayed graphically in Fig. 3: for  $p \leq p_c = r$  no infection survives ( $\phi^* \equiv 0$ ) while for  $p > p_c = r$ , a stable, finite fraction of the population will become infected ( $\phi^* > 0$ ). In the language of dynamical systems theory, the epidemic threshold is a transcritical bifurcation (Strogatz, 1994) which lies at the intersection of two fixed point curves, where the stable curve changes to unstable at the intersection and vice versa (in statistical mechanics, such behavior is referred to as a second-order or continuous phase transition, where  $(p_c, 0)$  is called the critical point (Goldenfeld, 1992)). For the choice of parameters above, one branch of the transcritical bifurcation is the  $p$ -axis which is stable to the left of  $p_c$  and unstable to

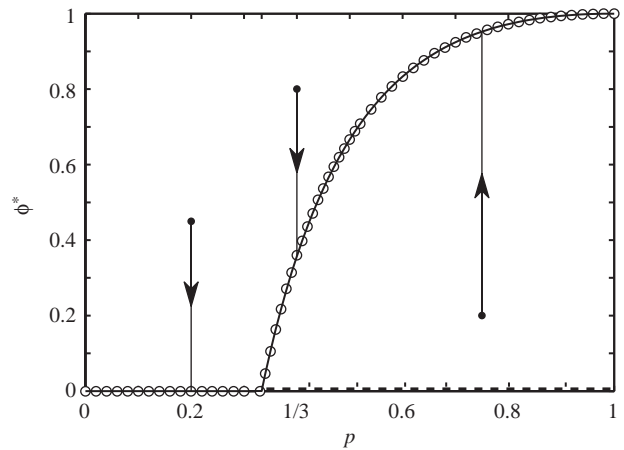


Fig. 3. Fixed point curves for an example epidemic threshold model. The stable fixed point curves shown are from simulation (circles) and Eq. (6) (line) for  $d^* = 1$ ,  $T = 2$ , and  $\tau = 1/r - 1 = 1$ . Unstable fixed points are indicated by the dashed line. The trajectories of three initial conditions are shown to illustrate how the level of an epidemic evolves for different values of  $p$ . The epidemic threshold  $p_c = \frac{1}{3}$  is as predicted by  $p_c = (T + \tau)^{-1}$ , Eq. (11). For the simulation, the population size  $N = 10^5$ , the number of time steps  $N_t = 10^3$ , each data point represents the average value of  $\phi^*(p)$  over the last  $N_s = 100$  time steps, and the initial condition is that all are infected.

the right (imagining the unphysical extension of  $\phi^*$  to  $\phi^* < 0$ , the other branch may be seen to extend from below the  $p$ -axis where it is unstable to above the  $p$ -axis where it is stable).

We observe this kind of bifurcation structure (i.e. where the rising branch has positive slope and comprises stable equilibria) to be a robust feature with respect to a range of parameter choices, and we classify all such models as epidemic threshold models. As we show below, the same qualitative equilibrium behavior is present in all homogeneous models with dose thresholds set at  $d^* = 1$ , even with arbitrary memory  $T \geq 1$  and recovery rate  $r \leq 1$ . While some details of the dynamics do change as  $T$  and  $r$  are varied, the existence of a single transcritical bifurcation depends only on the assumption that all individuals exhibit “trivial thresholds:”  $g(d^*) = \delta(d^* - 1)$ .

First, allowing memory to be arbitrary ( $T > 1$ ) but keeping  $r = 1$ , we observe that the individuals who are infected at some time  $t$  are necessarily those who have experienced at least one infectious event in the preceding  $T$  time steps; thus we obtain the following implicit equation for  $\phi^*$ :

$$\phi^* = 1 - (1 - p\phi^*)^T. \quad (6)$$

As with the  $T = 1$  case above, the equilibrium behavior exhibits a continuous phase transition and therefore also falls into the epidemic threshold class. While we can no longer find a general, closed-form solution for  $\phi^*$ , Eq. (6) can be rearranged to give  $p$  as an explicit

function of  $\phi^*$ :

$$p = \phi^{*-1}[1 - (1 - \phi^*)^{1/T}]. \tag{7}$$

Taking the limit of  $\phi^* \rightarrow 0$  in either Eq. (6) or Eq. (7), we find  $p_c = 1/T$ . Thus, receiving at least one exposure from the last  $T$  contacts is analogous to the zero memory ( $T = 1$ ), variable  $r$  case where recovery occurs on a time-scale  $1/r \simeq T$ .

Next, we also allow the recovery rate to be arbitrary ( $r < 1$ ). To find the fixed point curves, we modify Eq. (6) to account for the fraction of individuals that have not experienced a single exposure for at least the last  $T$  time steps but have not recovered from a previous infection. We first write down the probability that an individual last experienced an infectious event  $m$  time steps ago and has not yet recovered. Denoting the sequence of a positive unit dose followed by  $m$  0's as  $H_{m+1}$ , we have

$$P(\text{infected}|H_{m+1}) = p\phi^*(1 - p\phi^*)^m(1 - r)^{m-T+1}. \tag{8}$$

The first term on the right-hand side of the Eq. (8) is the probability of a successful exposure; the second term is the probability of experiencing no successful exposures in the subsequent  $m$  time steps; and the final term is the probability that, once the memory of the initial single exposure has been lost after  $T$  time steps, the individual remains infected. Since we are only concerned with individuals who have ‘forgotten’ the source of the infection, we have  $m \geq T$ . Summing over  $m$ , we obtain the total probability that an individual was infected at least  $T$  time steps ago and has not yet recovered:

$$\sum_{m=T}^{\infty} P(\text{infected}|H_{m+1}) = \frac{p\phi^*(1 - p\phi^*)^T(1 - r)}{1 - (1 - p\phi^*)(1 - r)}. \tag{9}$$

Adding this fraction to the right-hand side of Eq. (6) then gives

$$\begin{aligned} \phi^* &= 1 - (1 - p\phi^*)^T \left[ 1 - \frac{p\phi^*(1 - r)}{1 - (1 - p\phi^*)(1 - r)} \right] \\ &= 1 - \frac{r(1 - p\phi^*)^T}{1 - (1 - p\phi^*)(1 - r)}. \end{aligned} \tag{10}$$

Fig. 3 shows a comparison between the above equation and simulation results for  $T = 2$ ,  $r = \frac{1}{2}$ , and  $\bar{d}^* = 1$ .

Taking the limit of small  $\phi^*$  in Eq. (6) we find the epidemic threshold to be

$$p_c = \frac{r}{1 + r(T - 1)} = \frac{1}{T + 1/r - 1}. \tag{11}$$

Checking the special cases of the preceding calculations, we find  $p_c = r$  when  $T = 1$  and  $p_c = 1/T$  when  $r = 1$ . Denoting the mean time to recovery of an infected, isolated individual by  $\tau$ , we observe  $\tau = 1/r - 1$  and Eq. (11) becomes

$$p_c = \frac{1}{T + \tau}. \tag{12}$$

The time-scales  $T$  and  $\tau$  can thus both be thought of as corresponding to two different kinds of memory, the sum of which—the total characteristic time-scale of memory in the model—determines the position  $p_c$  of the epidemic threshold. Qualitatively, therefore, all homogeneous models that possess trivial individual thresholds exhibit the same kind of equilibrium dynamics. Varying the thresholds, however, produces equilibrium behavior of a quite distinct nature, as we show in the next section.

### 3.2. Critical mass models

When individuals of a homogeneous population require more than one exposure to become infected—that is, when  $g(\bar{d}^*) = \delta(\bar{d}^* - \bar{d}^*)$  and  $\bar{d}^* > 1$ —the closed-form expression for the fixed points  $\phi^*$  in the case of  $r = 1$  is

$$\phi^* = \sum_{i=\bar{d}^*}^T \binom{T}{i} (p\phi^*)^i (1 - p\phi^*)^{T-i}. \tag{13}$$

We now observe a fundamentally different behavior of the model: the transcritical bifurcation that is characteristic of epidemic threshold models is absent and is replaced by a saddle-node bifurcation (Strogatz, 1994), an example of which is illustrated in Fig. 4. We call models whose equilibrium states are determined in this manner *critical mass models* because, as indicated by the arrows in Fig. 4, an epidemic will not spread from an infinitesimal initial outbreak, requiring instead a finite fraction of the population, or ‘critical mass’ (Schelling, 1978), to be infected initially. Saddle-node bifurcations

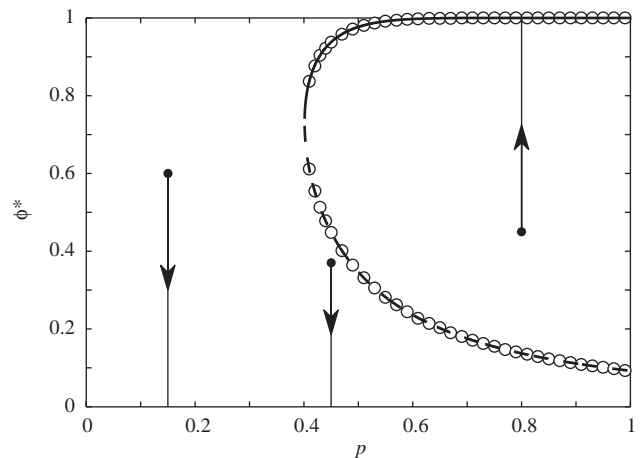


Fig. 4. Comparison between simulation (circles) and theory (lines, Eq. (13)) for  $T = 12$ ,  $\bar{d}^* = 3$ , and  $r = 1$ . For  $\bar{d}^* > 1$ , the homogeneous SIS contagion model exhibits a saddle-node bifurcation. Shown are the non-zero stable and unstable points. For the theoretical curves, solid lines represent stable points and dashed ones unstable points. All points on the line  $\phi^* = 0$  are stable points. The arrows show trajectories of the system for three example initial conditions. Simulations details are as per Fig. 3. The location of the unstable fixed point curve is determined by binary search.

(or backwards bifurcations) have also been observed in a number of unrelated epidemiological multi-group models, arising from differences between groups and inter-group contact rates (Haderler and Castillo-Chávez, 1995; Haderler and van den Driessche, 1997; Dushoff et al., 1998; Kribs-Zaleta and Velasco-Hernandez, 2000; Greenhalgh et al., 2000; Kribs-Zaleta, 2002).

Although it is not generally possible to obtain an expression for  $\phi^*(p)$  from Eq. (13), we are able to write down a closed-form expression involving the position of the saddle-node bifurcation  $(p_b, \phi_b^*)$  (see Appendix A for details):

$$0 = \sum_{i=\bar{d}^*}^T \binom{T}{i} z^{i-\bar{d}^*} (1-z)^{T-i-1} [i-1-z(T-1)], \quad (14)$$

where  $z = p_b \phi_b^*$ . Using Eq. (14), we solve for  $z$  by standard numerical means and then use Eq. (13) to obtain  $\phi_b^*$  and hence  $p_b$ . Fig. 5 shows positions of saddle-node bifurcation points computed for a range of values of  $T$  and  $\bar{d}^*$ . In all cases, as  $\bar{d}^*$  increases, the bifurcation point moves upward and to the right of the fixed point diagram. For small values of  $\bar{d}^*$  and  $T$ , we are able to determine  $(p_b, \phi_b^*)$  exactly. For example, for  $\bar{d}^* = 2$  and  $T = 3$ , we find that  $p_b = \frac{8}{9}$  and  $\phi_b^* = \frac{27}{32}$  and that the bifurcation is parabolic.

As with epidemic threshold models, the analysis for critical mass dynamics can be generalized to include a non-trivial recovery rate  $r < 1$ . Now, however, we are unable to find a closed-form expression for the fixed points for general  $T$  and  $\bar{d}^*$ . The difficulty in making such a computation lies in finding an expression for the number

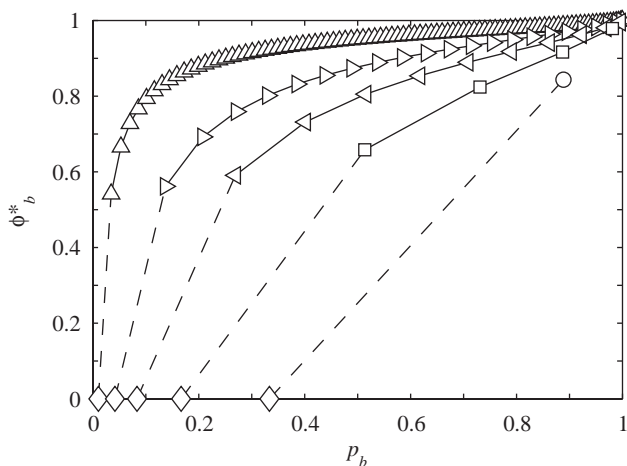


Fig. 5. Bifurcation points for  $\bar{d}^* > 1$ ,  $\phi^* > 1$ , and  $r = 1$ , determined numerically using Eqs. (13) and (14). The data shown are for  $T = 3$  ( $\circ$ ),  $T = 6$  ( $\square$ ),  $T = 12$  ( $\triangleleft$ ),  $T = 24$  ( $\triangleright$ ), and  $T = 96$  ( $\triangle$ ). The solid lines guide the eye for the range  $\bar{d}^* = 2$  to  $T - 1$  with  $\bar{d}^*$  increasing with  $p_b$ . The dashed lines connect to the transcritical bifurcation points ( $\diamond$ ) observed for  $\bar{d}^* = 1$ . Note that no bifurcations occur when  $\bar{d}^* = T > 1$  (the sole fixed point is  $\phi^* = 1$  when  $p = 1$ ) and when  $\bar{d}^* = T = 1$  (the fixed points lie along the line  $p = 1$  and  $0 \leq \phi \leq 1$ ).

of individuals whose dose counts are below threshold but are still infected since they have not yet recovered. For the  $\bar{d}^* = 1$  case, this was straightforward since the only way to stay below threshold was to experience a sequence of null exposures. Nevertheless, for  $r < 1$ , we can formally modify the expression for  $\phi^*$  given in Eq. (13):

$$\phi^* = \Gamma(p, \phi^*; r, T) + \sum_{i=\bar{d}^*}^T \binom{T}{i} (p\phi^*)^i (1-p\phi^*)^{T-i}, \quad (15)$$

where the additional term  $\Gamma(p, \phi^*; r)$  accounts for the proportion of below threshold individuals who have not yet recovered. For small values of  $T$  and  $\bar{d}^*$ , exact expressions for  $\Gamma(p, \phi^*; r)$  can be derived and then  $\phi^*$  can be solved for numerically. In Appendix B, we consider two special cases,  $\bar{d}^* = 2$  for  $T = 2$  and 3, that illustrate the process of constructing expressions for  $\Gamma(p, \phi^*; r)$ . These results allow us to explore the movement of the fixed points with decreasing  $r$ . Figs. 6A and B show

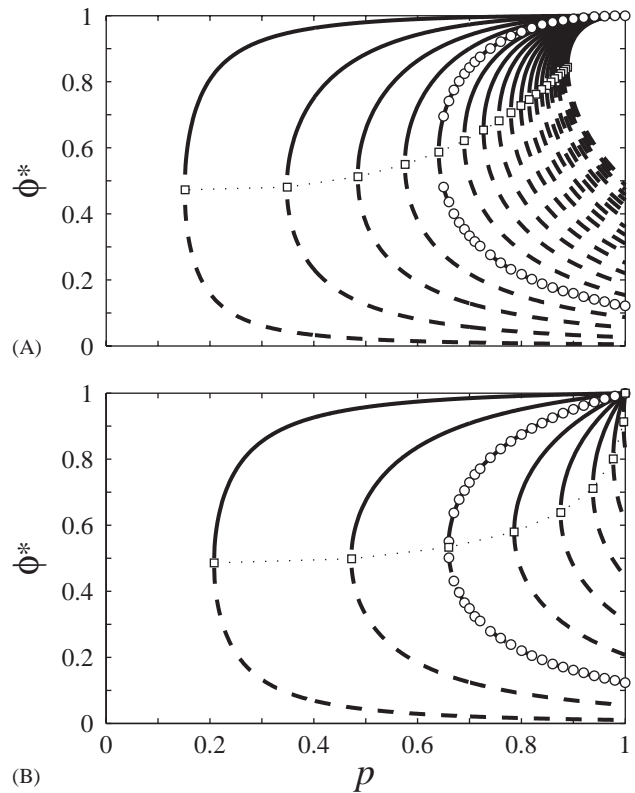


Fig. 6. (A) Theory versus simulation for  $T = 3$ ,  $\bar{d}^* = 2$ , and varying  $r$ . Solid lines represent theoretically derived stable fixed points, dashed lines represent unstable ones, and the squares indicate bifurcation points (using Eqs. (15) and (46)). From left-to-right, we have  $r = 0.01, 0.05, 0.10, 0.15, \dots, 1.00$ . The circles correspond to simulation data for the case  $r = 0.2$  and matches the theoretical curves. (B) Theoretical fixed point curves for  $T = 2$ ,  $\bar{d}^* = 2$ , and varying  $r$ . Bifurcations appear for  $r < 0.3820 \pm 0.0001$ . Otherwise,  $(p, \phi^*) = (1, 1)$  is the only fixed point of the system with  $\phi^* > 0$ . The values of  $r$  are as per the  $T = 3$  case in A. Circles correspond to simulation results for  $r = 0.1$ . For all values of  $r$ , both of the above systems possess a line of stable fixed points described by  $0 \leq p < 1$  and  $\phi^* = 0$ , i.e. the  $p$ -axis.

comparisons over a range of  $r$  between the solutions of Eq. (15) and simulations for  $T = 2$  and  $3$ , respectively, confirming that the agreement is excellent.

#### 4. Heterogeneous SIS contagion models

##### 4.1. Distributions of doses and thresholds

In real populations, both in the context of infectious diseases and also social influences, individuals evidently exhibit varying levels of susceptibility. Furthermore, contacts between infected and susceptible individuals can result in effective exposures of variable size, depending on the individuals in question, as well as the nature of their relationship and the circumstances of the contact (duration, proximity, etc.). Our homogeneity assumption of the preceding section is therefore unlikely to be justified in any real application. Furthermore, as we show below, the more general case of heterogeneous populations, while preserving much of the structure of the homogeneous case, also yields a new class of dynamical behavior; thus the inclusion of heterogeneity not only makes the model more realistic, but also provides additional qualitative insight.

While in principle all parameters in the model could be assumed to vary across the population, we focus here on two parameters—individual threshold  $d^*$  and dose size  $d$ —which when generalized to stochastic variables, embody the variations in individual susceptibility and contacts that we wish to capture. We implement these two sources of heterogeneity as follows. In the case of thresholds, each individual is assigned a threshold drawn randomly from a specified probability distribution  $g(d^*)$  at  $t = 0$  which remains fixed for all  $t$ . In effect, this assumption implies that individual characteristics remain roughly invariant on the time-scale of the dynamics, rather than varying from moment to moment. By contrast, in order to capture the unpredictability of circumstance, we assign dose sizes stochastically, according to the distribution  $f(d)$ , independent both of time and the particular individuals between whom the contact occurs.

Again considering the more tractable  $r = 1$  version of the SIS model, we first examine the effect of allowing dose size  $d$  to vary while holding thresholds  $d^*$  fixed across the population. In the homogeneous case,  $k$  exposures of a susceptible to infected individuals resulted in a dose count of  $k$ , but the result can be more complicated when  $d$  is allowed to vary continuously. Note that  $d^*$  also no longer need be an integer. First, we calculate the probability that a threshold will be exceeded by  $k$  doses. As the distribution of dose size is now some arbitrary function  $f$ , we have that the probability distribution of the sum of  $k$  doses is given by

the  $k$ -fold convolution

$$P\left(\sum_{j=1}^k d_j = d\right) = f * f * \dots * f(d) = f^{k*}(d). \tag{16}$$

The probability of exceeding  $d^*$  is then

$$P\left(\sum_{j=1}^k d_j \geq d^*\right) = \int_{d=d^*}^{\infty} dd f^{k*}(d). \tag{17}$$

Since, for  $r = 1$ , individuals are only infected when their dose count exceeds their threshold, we have that the steady-state fraction infected is given by

$$\phi^* = \sum_{k=1}^T \binom{T}{k} (p\phi^*)^k (1 - p\phi^*)^{T-k} \int_{d=d^*}^{\infty} dd f^{k*}, \tag{18}$$

where we have averaged over all possible ways an individual may experience  $1 \leq k \leq T$  exposures in  $T$  interactions.

Next, in order to account for any variation  $d^*$ , we must incorporate another layer of averaging into Eq. (18) as follows:

$$\begin{aligned} \phi^* &= \int_0^{\infty} dd^* g(d^*) \sum_{k=1}^T \binom{T}{k} (p\phi^*)^k (1 - p\phi^*)^{T-k} \\ &\quad \text{Times} \int_{d=d^*}^{\infty} dd f^{k*}(d) \\ &= \sum_{k=1}^T \binom{T}{k} (p\phi^*)^k (1 - p\phi^*)^{T-k} P_k, \end{aligned} \tag{19}$$

where  $P_k$  is defined by Eq. (3). An important insight that can be derived immediately from Eq. (19) is that all information concerning the distributions of dose sizes and thresholds is expressed via the  $\{P_k\}$ ; hence the details of the functions  $f$  and  $g$  are largely unimportant. In other words, many pairs of  $f$  and  $g$  can be constructed to give rise to the same  $\{P_k\}$  and hence the same fixed points. For example, any desired  $\{P_k\}$  can be realized by a uniform distribution of unit doses  $f(d) = \delta(d - 1)$ , along with a discrete distribution of thresholds

$$g(d^*) = P_1 \delta(d^* - 1) + \sum_{k=1}^{T-1} (P_{k+1} - P_k) \delta(d^* - k - 1). \tag{20}$$

We observe that for the homogeneous case where doses and thresholds are fixed at  $1$  and  $\bar{d}^*$ , respectively [i.e.  $f(d) = \delta(d - 1)$  and  $g(d^*) = \delta(d^* - \bar{d}^*)$ ], we have  $f^{k*}(d) = \delta(d - k)$  and the expression for  $P_k$  [Eq. (3)] simplifies to  $P_k = 0$  if  $k < \bar{d}^*$  and  $P_k = 1$  if  $k \geq \bar{d}^*$ . Substituting these conditions into Eq. (19), we recover our previous expression Eq. (13) for the  $r = 1$  homogeneous case.



### 4.2. Universal classes of contagion

In the homogeneous version of the model, we determined the existence of two classes of dynamics—epidemic threshold and critical mass—with the former arising whenever  $\bar{d}^* = 1$ , and the latter when  $\bar{d}^* \geq 2$ . In other words, in homogeneous populations, the condition for differentiating between one class of behavior and another is a discrete one. Once heterogeneity is introduced, however, we observe a smooth transition between epidemic threshold and critical mass models governed by a continuous adjustment of the distributions  $f$  and  $d$  (or equivalently the  $\{P_k\}$ ). One consequence of this now-continuous transition is the appearance of an intermediate class of models, which we call *vanishing critical mass* (see Fig. 7). As with pure critical mass models, this new class is characterized by a saddle-node bifurcation, but now the lower unstable branch of fixed points crosses the  $p$ -axis at  $p_c$ ; in other words, the required critical mass “vanishes.” The collision of the unstable branch of the saddle node bifurcation with the horizontal axis also effectively reintroduces a transcritical bifurcation, in the manner of epidemic threshold models. However, the transcritical bifurcation is different from the one observed in epidemic threshold models because the rising branch of the bifurcation has negative, rather than positive, slope and comprises unstable, rather than stable, fixed points. Vanishing critical mass dynamics are therefore qualitatively distinct from both previously identified classes of behavior, exhibiting important properties of each: for  $p < p_c$ , they behave like critical mass models; and for  $p > p_c$ , they behave like epidemic threshold models, in the sense that an infinitesimal initial seed can spread.

Our generalized model therefore exhibits behavior that falls into one of only three universal classes: class I (epidemic threshold), class II (vanishing critical mass), and class III (critical mass). As we show below, more complicated fixed point curves exist (i.e. curves possessing two or more saddle-node bifurcation points) but nevertheless belong to one of these three universal classes, since together they include all possible behaviors of the fixed point curves near  $p = p_c$ .

In addition to identifying three universal classes of behavior, it is also possible to specify the conditions that govern into which class any particular choice of model parameters will fall. For  $r = 1$ , we can calculate when the transitions between universal classes occur in terms of the parameters of interest; i.e. the  $\{P_k\}$  (viz.  $f$  and  $g$ ) and  $T$ . This exercise amounts to locating the transcritical bifurcation and determining when it collides with the saddle-node bifurcation. To locate the transcritical bifurcation, we examine the fixed point equation as  $\phi^* \rightarrow 0$ . Since, from Eq. (19),

$$\phi^* = Tp\phi^*P_1 + O(\phi^{*2}), \tag{21}$$

we have

$$p_c = \frac{1}{TP_1}. \tag{22}$$

The position of the transcritical bifurcation is therefore determined by the memory length  $T$  and the fraction of individuals who are typically infected by one exposure,  $P_1$ . We see immediately that the transition between classes II and III contagion models occurs when  $p_c = 1$ ; i.e. when

$$P_1 = 1/T. \tag{23}$$

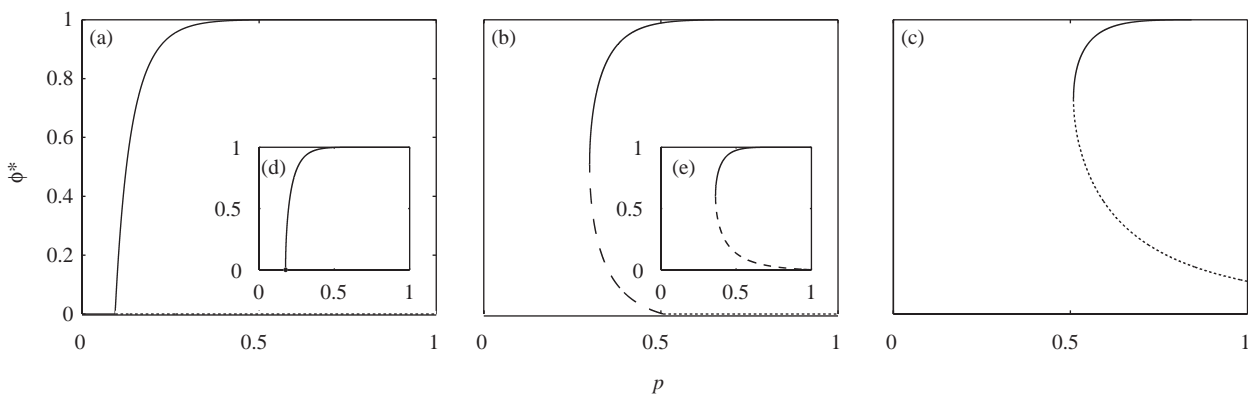


Fig. 7. Fixed point curves for the three main classes of contagion models produced by the model, along with transitions between classes. All curves are determined by numerical solution of Eq. (19) with an error tolerance of  $10^{-10}$ . Here,  $T = 12$ , individual doses are lognormally distributed with mean 1 and a variance of 0.6 for the underlying normal distribution, and thresholds are homogeneously distributed with  $g(d^*) = \delta(d^* - \bar{d}^*)$ . The main plots correspond to (a)  $\bar{d}^* = 0.4$  (class I), (b)  $\bar{d}^* = 1.5$  (class II), and (c)  $\bar{d}^* = 3$  (class III). The inset plots (d) and (e), respectively, show fixed point curves for models at the classes I–II transition (i.e. when  $P_1 = P_2/2$  which occurs when  $\bar{d}^* = 0.8600\dots$ ) and the classes II–III transition (i.e. when  $p_c = 1$  which occurs when  $\bar{d}^* = 1.9100\dots$ ). In all plots except (e), the intersection between the fixed point curve and the  $p$ -axis is a transcritical bifurcation. In plots (b), (c), and (e), the second bifurcation is a saddle-node bifurcation.

Recalling the homogeneous case, we see that when  $\bar{d}^* = d = 1$ ,  $P_1 = 1$ , giving  $p_c = 1/T$  as before. It necessarily follows that when  $\bar{d}^* > d = 1$ ,  $P_1 = 0$  and therefore  $p_c = \infty$ , thus confirming our earlier finding that the homogeneous model with  $\bar{d}^* > 1$  is always in the pure critical mass class (i.e. the lower, unstable fixed point curve of the saddle node bifurcation must have  $\phi^* > 0$  for all  $p_b \leq p \leq 1$  since it only reaches  $\phi^* = 0$  in the limit  $p \rightarrow \infty$ ). We also see that technically all models possess a transcritical bifurcation somewhere along the  $p$ -axis, even though it may be located beyond  $p_c > 1$  (when  $P_1 = 0$ , it lies at  $p_c = \infty$ ).

In order to locate the transition between classes I and II, we determine when the transcritical and saddle-node bifurcations are coincident, i.e. when  $d\phi^*/dp = \infty$  at  $(p, \phi^*) = (p_c, 0)$ . In other words, we calculate when the fixed point curve emanating from  $p = p_c$  is at the transition between having a large positive slope (class I) and a large negative slope (class II). We find the condition for this first transition is

$$P_1 = P_2/2, \tag{24}$$

where details of this calculation are provided in Appendix C.

The condition of Eq. (24) is a statement of linearity in the  $\{P_k\}$ , though importantly only for  $k = 1$  and 2. Providing  $p_c < 1$ , if  $P_1 > P_2/2$  (i.e. sublinearity holds) a contagion model is class I whereas if  $P_1 < P_2/2$  (i.e. superlinearity holds) it is class II. This condition means that class II contagion models arise when, on average, two doses are more than twice as likely as one dose to cause infection.

When this condition is satisfied exactly, the system's phase transition is a continuous one as per those of class I, but of a different universality class: class I systems exhibit a linear scaling near the critical point whereas the scaling when  $P_1 = P_2/2$  is  $\phi^* \propto (p - p_c)^{1/2}$  (in Appendix C, we show that a sequence of increasingly specific exceptions to this scaling exist depending on the extent of linearity present in the  $\{P_k\}$ ).

Thus, for  $r = 1$ , the condition that determines whether a given system is described by an epidemic threshold model, a critical mass model, or a member of the intermediate class of vanishing critical mass models, depends only on  $T$ ,  $P_1$ , and  $P_2$ —a surprising result given that Eq. (19) clearly depends on all the  $\{P_k\}$ . A summary of the three basic system types, the transitions between them, and the accompanying conditions is given in Table 2.

For  $r < 1$ , the position of the transcritical bifurcation, Eq. (25), generalizes in the same manner as Eq. (12). We find

$$p_c = \frac{1}{P_1(T + \tau)}, \tag{25}$$

Table 2

Summary of basic states of the  $r = 1$ , heterogeneous version of model along with the corresponding parameter ranges

I: Epidemic threshold	$P_1 > P_2/2$ and $P_1 > 1/T$
I–II transition	$P_1 = P_2/2$ and $P_1 > 1/T$
II: Partial critical mass	$P_1 < P_2/2$ and $P_1 > 1/T$
II–III transition	$P_1 < P_2/2$ and $P_1 = 1/T$
Pure critical mass	$P_1 < P_2/2$ and $P_1 < 1/T$

The  $P_k$  quantities depend on the distributions of dose size and individual thresholds, and are given in Eq. (3).

where we recall that  $\tau = 1/r - 1$ . As we will see in Section 5, the above statement is also true for  $\rho < 1$  and thus stands as a completely general result for the model. We therefore have a condition for the transition between classes II and III contagion models for a given  $T$  and  $r$ , analogous to that for the  $\rho = 1$  case in Eq. (23):  $P_1 = 1/(T + \tau)$ . For  $r < 1$ , the condition for the transition between classes I and II contagion models is more complicated both in derivation and form. We observe that as  $r$  is decreased, class III models must at some point transition to class II models and class II models will eventually become class I models, where we can determine the former transition by setting  $p_c = 1$  in Eq. (25) and solving for  $r$ :

$$r = \frac{P_1}{1 + P_1 - P_1 T}. \tag{26}$$

In Eq. (26) we have assumed  $P_1 > 0$  and  $P_1 T < 1$ , since the  $r = 1$  limit is, by assumption, a class III contagion model. Fig. 8 presents an example of a model transitioning from class III through II to I.

### 4.3. Composite classes of dynamics

Although our main results (three universal classes of behavior and the conditions that govern the transitions between the three) involve examples of contagion models with at most two bifurcations, other more complicated kinds of equilibrium behavior are possible. Fig. 9 shows four examples of what can happen for particular distributions of  $\bar{d}^*$  across the population. In each example,  $T = 20$  and  $r = 1$ , all doses are of unit size, and the population is divided into either two or three subpopulations with distinct values of  $\bar{d}^*$ . The main features of each system are captured by the number and locations of the saddle-node bifurcations. As was the case for the homogeneous model [see Eq. (14)], we are able to find an expression for  $z = p_b \phi_b^*$ :

$$0 = \sum_{k=1}^T \binom{T}{k} P_k z^{k-2} (1-z)^{T-k-1} [k-1 - z(T-1)], \tag{27}$$

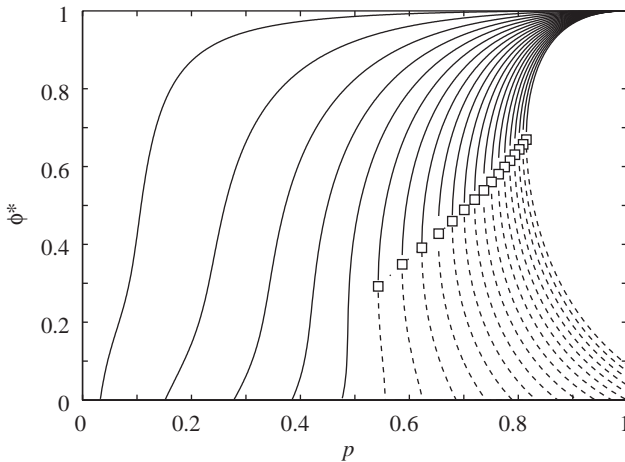


Fig. 8. Example of transitions between all three classes of contagion models occurring as the probability of recovery  $r$  changes. Shown are theoretically derived fixed point curves for a range of  $r$  for  $T = 3$  and a heterogeneous population with threshold distribution  $g(d^*) = 0.3\delta(d^* - 1) + 0.7\delta(d^* - 2)$ . From left-to-right, the curves correspond to  $r = 0.01, 0.05, 0.10, 0.15, \dots, 1.00$ . Solid lines indicate stable fixed points, dashed lines unstable ones, and the squares mark saddle-node bifurcation points. For  $r > 0.75$ , the system belongs to class III (see Eq. (26)). Fixed points lying on the  $p$ -axis are not indicated in the plot. For each instance of the model, points on the  $p$ -axis to the left of  $p_c$  (the intercept of the non-zero fixed point curve with the  $p$ -axis) are stable and points to the right are unstable. The expression used here to determine the fixed points is formed by appropriately weighting a combination of Eqs. (6), (15), and (46).

where the details of this calculation are provided in Appendix A. In principle, Eq. (27) could be analysed to deduce which  $\{P_k\}$  (and hence which  $f$  and  $g$ ) lead to what combination of bifurcations. While substantially more complicated, and thus beyond the scope of this paper, such a classification scheme would be a natural extension of our present delineation of the model into three universal classes of contagion models, based on the behavior near  $p = p_c$ . One simple observation, however, is that Eq. (27) is a polynomial of order  $T - 2$  and hence a maximum of  $T - 2$  saddle-node bifurcations may exist. From our investigations this outcome seems unlikely as nearby bifurcations tend to combine with or overwhelm one another; in other words, subpopulations with sufficiently distinct  $d^*$  are required to produce systems with multiple saddle-node bifurcations. Furthermore, the distribution of dose sizes  $f$  seems unlikely to be multimodal for real contagious influences or entities, and the way it enters into the calculation of the  $\{P_k\}$  [Eq. (3)] reduces its effect in producing complicated systems. Thus, the number of distinct bifurcations is limited and strong multimodality in the threshold distribution  $g$  appears to be the main mechanism for producing systems with more than one saddle-node bifurcation. As a first step in this extended analysis of the model, we derive in Appendix A the

condition for the appearance of two saddle-node bifurcations (i.e. one forward and one backward).

## 5. SIRS and SIR contagion models

As mentioned in Section 2, the SIS class of behavior that we have analysed exclusively up to now is a special case of the general contagion process as it assumes that recovered individuals instantly become re-susceptible. This assumption renders the SIS case particularly tractable, and we have taken advantage of this fact in the preceding sections to make some headway in understanding the full range of equilibrium behavior of the model. However, it remains the case that very few, if any, infectious diseases could be considered to obey true SIS-type dynamics, as almost all recovery from infection tends to be associated with some finite period of immunity. Any purportedly “general” model of contagion ought therefore to be analysed in a wider domain of the associated removal period, and any corresponding classes of behavior labeled “universal” ought to withstand the introduction of at least some period of immunity to re-infection. Thus motivated, we now extend our previous analysis to systems where individuals experience temporary (SIRS,  $0 < \rho < 1$ ) or permanent removal (SIR,  $\rho = 0$ ). We present some preliminary results for each of these cases in turn, relying now exclusively on numerical simulation.

For SIRS contagion, we observe that the position of the transcritical bifurcation  $(p_c, 0)$  does not change as  $\rho$  is reduced from 1; however, all non-zero fixed points move in the positive  $p$  direction. Because they remain in the removed state for a longer time, individuals in systems with lower  $\rho$  spend relatively less time infected than those in systems with higher  $\rho$  (this is in contrast to the effect of reducing  $r$ , which prolongs the time individuals are infected, thereby causing fixed points to move in the negative  $p$  direction). Thus, contagion models belonging to classes I and III in the  $\rho = 1$  special case remain in their respective classes as  $\rho$  decreases. However, as shown in Fig. 10, class II models will transition to class I for some  $\rho < 1$ .

For the SIR version of the model, recovered individuals cannot return to the susceptible state, and no fraction of the population remains infected in the infinite time limit; hence we can no longer speak of non-zero fixed point curves, and the fraction of individuals infected and recovered,  $\phi(t)$  and  $R(t)$ , become the relevant objects of study. Some defining quantities are then the maximum fraction of the population infected at any one time,  $\max_t \phi(t)$ , the fraction eventually infected,  $1 - \phi(\infty) = R(\infty)$ , and the *relaxation time* required for the epidemic to die out. We focus on the latter here which we denote by  $t_{\max}$ .

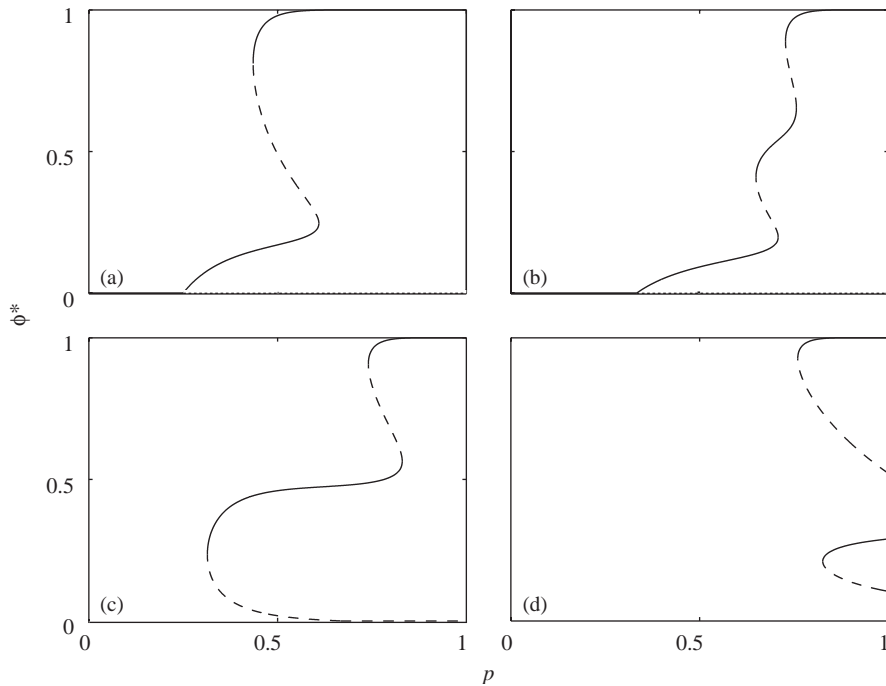


Fig. 9. Fixed point curves involving multiple bifurcations. In all four examples,  $T = 20$ ,  $r = 1$ , and dose size is uniformly held at unity. The different curves are obtained by adjusting the threshold distribution  $g$  which in turn leads to changes in the  $\{P_k\}$  [see Eq. (3)]. The curves correspond to (a)  $g(d) = 0.2\delta(d - 1) + 0.8\delta(d - 6)$ ; (b)  $g(d) = 0.15\delta(d - 1) + 0.4\delta(d - 5) + 0.45\delta(d - 12)$ ; (c)  $g(d) = 0.075\delta(d - 1) + 0.4\delta(d - 2) + 0.525\delta(d - 12)$ ; and (d)  $g(d) = 0.3\delta(d - 3) + 0.7\delta(d - 12)$ . Note that example (d) consists of two separate fixed point curves. The curves were found numerically by solving for  $p(\phi^*)$  using Eq. (19). Solid and dashed lines indicate stable and unstable fixed points, respectively.

We observe that when individuals possess a memory of doses (i.e.  $T > 1$ ),  $t_{\max}$  diverges as  $p \rightarrow 1$ . Fig. 11 shows  $t_{\max}(p)$  for four sample systems with  $T = 3, 4, 5$ , and 6. We find the divergence of  $t_{\max}$  near  $p = 1$  to be well approximated by

$$t_{\max} \propto (1 - p)^{-(T-1)}, \tag{28}$$

where fits are shown in the inset of Fig. 11. Eq. (28) also shows that for a fixed  $p$ , the relaxation time increases exponentially with length of memory  $T$  for all  $p$ , i.e.

$$t_{\max} \propto e^{\alpha T}, \tag{29}$$

where  $\alpha = -\ln(1 - p) > 0$ .

Thus, when  $\rho = 0$ , epidemics, while not ever achieving a non-zero steady state as in the  $\rho > 0$  case, can persist for (arbitrarily) long periods of time. The introduction of memory, which allows infected individuals to maintain their dose count above their threshold by repeatedly infecting each other, creates an SIR model with strikingly different behavior to the standard memoryless SIR model.

### 6. Concluding remarks

Our aim here has been to develop and analyse in detail a model of contagion that incorporates and

generalizes elements of contagion models from both the social sciences and epidemiology. A key feature of the current model is that interdependencies between successive exposures are introduced in a natural way by varying the length of memory that an individual maintains of past exposures. Contagion models incorporating memory correspond to standard notions of contagion in the social sciences (although these models rarely discuss the role of memory explicitly), while memoryless contagion models correspond to traditional models of disease spreading. Our model suggests, however, that in reality these two kinds of contagion models may not be entirely distinct. We venture the possibility that some infectious diseases may spread in a fashion similar to social contagion processes. For example, two exposures to an agent sufficiently close in time may infect an individual with higher probability than would be expected if the exposures acted as independent events. The response of some individuals to allergens where separate doses accumulate in the body may be an example of a disease with memory. Although allergies are not contagious, they demonstrate that such a phenomenon is biologically plausible—a possibility that has largely remained unexamined in the microorganismal dose–response literature (Haas, 2002).

The main result of our analysis is the identification of three universal classes of contagion dynamics, along

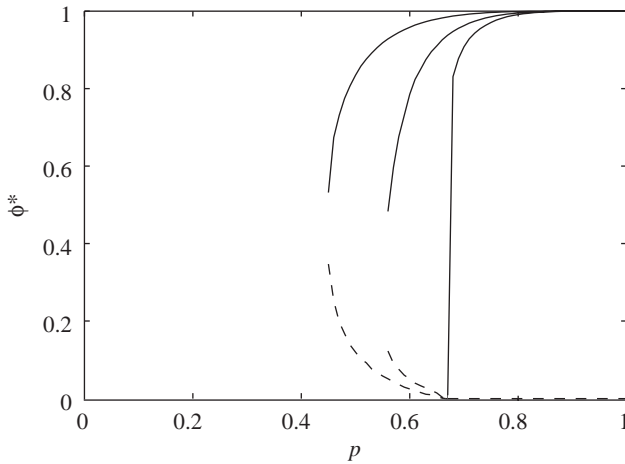


Fig. 10. The effect of reducing  $\rho$  (the probability of an immune individual becoming susceptible) for a contagion model that is class II when  $\rho = 1$ . Here,  $N = 10^5$ ,  $T = 6$ ,  $r = 1$ , dose sizes are fixed at unity, and  $g(d^*) = 0.25\delta(d^* - 1) + 0.75\delta(d^* - 2)$ . From left-to-right, the plots correspond to  $\rho = 1, 0.15$ , and  $0.09$ . In all cases,  $p_c$ , the position of the  $p$ -axis intercept, is independent of  $\rho$ . Apart from the intercept at  $p = p_c$  and the point  $(p, \phi^*) = (1, 1)$ , all other points of the curve move to the right as  $\rho$  decreases. Consequently, as  $\rho$  is reduced, all class II contagion models will at some point become members of class I. Contagion models in classes I and III remain unchanged in their nature. In finding the upper stable branch of  $\phi^*(p)$ , all individuals are initially infected. A binary search is then used to detect the position of the lower unstable branch.

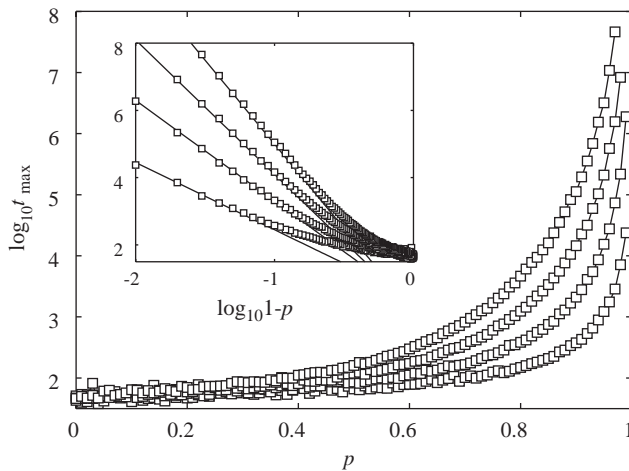


Fig. 11. Time taken for an initially universal infection to die out in the SIR case ( $\rho = 0$ ). Moving from bottom-to-top in both plots,  $T = 3, 4, 5$ , and  $6$ . The model parameters used here are  $d^* = 1$ , unit dose size,  $r = 0.2$ , and population size  $N = 10^4$ . For all systems,  $p = \phi = 1$  is a fixed point and hence  $t_{\max} = \infty$  at  $p = 1$ . The main plot shows the rapid increase in  $t_{\max}$  for  $p \rightarrow 1$ . The inset shows the behavior of  $t_{\max}$  as a function of  $(1 - p)$  as plotted on a double logarithmic scale. The lines have slopes of  $-(T - 1)$  indicating that infections are strongly persistent with  $t_{\max} \propto (1 - p)^{-(T-1)}$ .

with precise conditions for the transitions between these classes. Given the complexity of the model, these conditions are surprisingly simple—at least in the SIS

case ( $r = \rho = 1$ , see Table 2)—and present us with quantities such as  $T$ ,  $P_1$ , and  $P_2$  that may in principle be measurable for real epidemics. Furthermore, the dependence of the transition conditions only on  $P_1$ ,  $P_2$ , and  $T$ , rather than on the full details of the underlying distributions of thresholds ( $g(d^*)$ ) and doses ( $f(d)$ ), suggests a new and possibly useful level of abstraction for thinking about contagion processes; that is, measuring individuals in terms of their dose–response and characterizing a population in terms of its  $\{P_k\}$ .

For the more complicated and general cases of  $r < 1$  (finite recovery period) and  $\rho < 1$  (finite immunity period), we have confirmed that the same basic three-class structure persists, and determined the position of the transcritical bifurcation  $p_c$ , Eq. (25), that is one of the two quantities needed to specify the conditions for transitioning between classes. The other condition, derived from calculating the slope of the fixed point curve as it passes through the transcritical bifurcation, merits further attention. All of these conditions, however, are ultimately dependent on the distributions  $f$  and  $g$ . Our analysis of the model suggests, for example, that composite fixed point diagrams involving more than one saddle point node can only result from multimodality in  $g$ , the distribution of thresholds. Exactly how the details of these two distributions affect  $P_1$  and  $P_2$  would be worth further investigation.

Our model suggests that some epidemics may be prevented or enabled with slight changes in system parameters (if feasible). For example, knowing that a potentially contagious influence belongs to class II and that  $p$  is just below  $p_c$  would indicate that by increasing inherent infectiousness (i.e.  $p$ ) or by creating a sufficiently large enough base of infected individuals, the contagion could be kicked off with potentially dramatic results. Alternatively, by increasing  $r$  or reducing  $T$  or  $\rho$ , the possibility of undesirable epidemics may be reduced, as for all these adjustments fixed point curves are generally moved in the direction of higher values of  $p$ .

Finally, we have focused exclusively in this paper on a mean field analysis of the model, which is to say, we have made the standard assumption that individuals in a population mix uniformly and at random. A natural generalization would be to consider the model's behavior for a networked population of individuals. Other simulation possibilities would be to consider distributions of  $T$ ,  $p$ ,  $r$ , and  $\rho$ . Finally, additional technical investigation of this model would also include analysis of the closed-form expression for saddle-node bifurcations given in Eq. (27), and a derivation of analytic expressions for the model when  $\rho < 1$  for small  $T$  and  $d^*$ . We hope that our preliminary investigations into this interesting and reasonably general class of contagion models will stimulate other researchers to pursue some of these extensions.

**Acknowledgements**

We are grateful for discussions with Duncan Callaway, Daryl Daley, Charles Haas, and Matthew Salganik. This research was supported in part by the National Science Foundation (SES 0094162), the Office of Naval Research, Legg Mason Funds, and the James S. McDonnell Foundation.

**Appendix A. Conditions for existence of saddle-node bifurcation**

In Eq. (27), we provided a closed form expression for  $z = p_b \phi_b^*$  where  $(p_b, \phi_b^*)$  is the location of a saddle-node bifurcation. This expression pertains to the heterogeneous version of the model for  $r = \rho = 1$  [the homogeneous version is given in Eq. (14)]. As noted in the main text, this equation has up to  $T - 2$  solutions, depending on the form of the  $\{P_k\}$ . In this Appendix, we derive Eq. (27) and also find a criterion for the appearance of two saddle-node bifurcations. All calculations revolve around determining when the slope of the fixed point curve  $\phi^*(p)$  becomes infinite, or, equivalently, finding when  $dp/d\phi^* = 0$ .

Our starting point is Eq. (19), the general closed form expression for  $\phi^*$  as a function of  $p$ , from which we can calculate  $dp/d\phi^*$ . We rewrite Eq. (19) as

$$\begin{aligned} u(p, \phi^*) &= -\phi^* + \sum_{k=1}^T \binom{T}{k} P_k (p\phi^*)^k (1 - p\phi^*)^{T-k} \\ &= \phi^* \left[ -1 + p \sum_{k=1}^T \binom{T}{k} P_k (p\phi^*)^{k-1} (1 - p\phi^*)^{T-k} \right] \\ &= \phi^* U(p, \phi^*) \end{aligned} \tag{30}$$

with the requirement  $u(p, \phi^*) = 0$ .

We show that we can use the function  $U$  instead of  $u$  to find  $dp/d\phi^*$  when  $\phi^* \neq 0$ . Differentiating  $u(p, \phi^*) = \phi^* U(p, \phi^*) = 0$  with respect to  $\phi^*$ , we have

$$\frac{\partial u}{\partial \phi^*} + \frac{dp}{d\phi^*} \frac{\partial h}{\partial p} = U + \phi^* \frac{\partial U}{\partial \phi^*} + \phi^* \frac{dp}{d\phi^*} \frac{\partial U}{\partial p} = 0. \tag{31}$$

Again, since  $u(p, \phi^*) = \phi^* U(p, \phi^*) = 0$ , we have  $U(p, \phi^*) = 0$  when  $\phi^* \neq 0$ . When we also require  $dp/d\phi^* = 0$  (i.e. the chief condition for a saddle-node bifurcation point), Eq. (31) reduces to

$$\frac{\partial U}{\partial \phi^*} = 0 \tag{32}$$

and so we may find solutions of  $\partial U/\partial \phi^* = 0$  instead of  $\partial u/\partial \phi^* = 0$ . The benefit of making this observation is that we find  $\partial U/\partial \phi^*$  can be expressed in terms of a single variable ( $z = p_b \phi_b^*$ ), allowing for simpler analytical and numerical examination (recall that  $(p_b, \phi_b^*)$  denotes the position of a saddle-node bifurcation).

Returning to the definition of  $U$  given in Eq. (30), we find

$$\begin{aligned} \frac{\partial U}{\partial \phi^*} &= \sum_{k=1}^T \binom{T}{k} P_k p^k (k-1) (\phi^*)^{k-2} (1 - p\phi^*)^{T-k} \\ &\quad + \sum_{k=1}^T \binom{T}{k} P_k p^k (\phi^*)^{k-1} (T-k) (-p) \\ &\quad \times (1 - p\phi^*)^{T-k-1} + \frac{dp}{d\phi^*} (\dots). \end{aligned} \tag{33}$$

Since we require  $dp/d\phi^* = 0$ , the above simplifies to

$$\begin{aligned} \frac{\partial U}{\partial \phi^*} &= \sum_{k=1}^T \binom{T}{k} P_k p^k (\phi^*)^{k-2} (1 - p\phi^*)^{T-k-1} \\ &\quad \times [(k-1)(1 - p\phi^*) - p\phi^*(T-k)]. \end{aligned} \tag{34}$$

Setting  $\partial U/\partial \phi^* = 0$  and removing a factor of  $p^2$ , we find the positions of all saddle-node bifurcation points satisfy

$$0 = \sum_{k=1}^T \binom{T}{k} P_k z^{k-2} (1-z)^{T-k-1} [k-1 - z(T-1)], \tag{35}$$

where  $z = p_b \phi_b^*$ . Upon solving Eq. (35) for  $z$  (where for all non-trivial solutions, we require  $0 < z < 1$ ), Eq. (19) can then be used to find  $\phi_b^*$  (since it expresses  $\phi_b^*$  as a function of  $z$ ) and hence  $p_b$ .

In order to determine whether a bifurcation is forward or backward facing (by forward facing, we mean branches emanate from the bifurcation point in the direction of the positive  $p$ -axis), we need to compute  $d^2p/d\phi^{*2}$ , and examine its sign. (When  $d^2p/d\phi^{*2} = 0$ , two saddle-node bifurcations points are coincident, one forward and one backward facing.) If the  $\{P_k\}$  are parametrized in some fashion (i.e.  $f$  and/or  $g$  are parametrized), then we can determine the relevant parameter values at which pairs of bifurcations appear. We compute an expression for  $d^2p/d\phi^{*2}$  as follows.

Differentiating Eq. (31) with respect to  $\phi^*$ , we have

$$\frac{d^2u}{d\phi^{*2}} = 2 \frac{dU}{d\phi^*} + \phi^* \frac{d^2U}{d\phi^{*2}} = 0. \tag{36}$$

We already have  $dU/d\phi^* = \partial U/\partial \phi^* + dp/d\phi^* \partial U/\partial p = 0$ , and so Eq. (36) now gives

$$\frac{d^2U}{d\phi^{*2}} = 0. \tag{37}$$

Expanding this, we have

$$0 = \frac{\partial^2 U}{\partial \phi^{*2}} + \frac{dp}{d\phi^*} \frac{\partial^2 U}{\partial p \partial \phi^*} + \frac{d^2p}{d\phi^{*2}} \frac{\partial U}{\partial p} + \frac{dp}{d\phi^*} \frac{d}{dp} \frac{\partial U}{\partial p}. \tag{38}$$

The second and fourth terms on the right-hand side disappear since  $dp/d\phi^* = 0$ , leaving

$$0 = \frac{\partial^2 U}{\partial \phi^{*2}} + \frac{d^2 p}{d\phi^{*2}} \frac{\partial U}{\partial p}. \quad (39)$$

Upon rearrangement, we have

$$\frac{d^2 p}{d\phi^{*2}} = -\frac{\partial^2 U / \partial \phi^{*2}}{\partial U / \partial p}. \quad (40)$$

We first compute  $\partial U / \partial p$ . With  $U$  as defined in Eq. (30), we see that

$$\begin{aligned} \frac{\partial U(p, \phi^*)}{\partial p} &= \frac{\partial}{\partial p} \frac{1}{\phi^*} \sum_{k=1}^T \binom{T}{k} P_k (p\phi^*)^k (1 - p\phi^*)^{T-k} \\ &= \frac{\partial}{\partial p\phi^*} \sum_{k=1}^T \binom{T}{k} P_k (p\phi^*)^k (1 - p\phi^*)^{T-k} \\ &= \frac{\partial}{\partial z} \sum_{k=1}^T \binom{T}{k} P_k z^k (1 - z)^{T-k} \\ &= \sum_{k=1}^T \binom{T}{k} P_k z^k (1 - z)^{T-k-1} \\ &\quad \times [k(1 - z) - (T - k)z] \\ &= \sum_{k=1}^T \binom{T}{k} P_k z^k (1 - z)^{T-k-1} [k - Tz]. \end{aligned} \quad (41)$$

Using Eqs. (30) and (35) (i.e. that we are at a bifurcation point), the above yields

$$\frac{\partial U(p, \phi^*)}{\partial p} = \frac{1}{p}. \quad (42)$$

Using this result, Eq. (40) becomes

$$\frac{d^2 p}{d\phi^{*2}} = -p \frac{\partial^2 U}{\partial \phi^{*2}}. \quad (43)$$

Next, we find

$$\begin{aligned} \frac{\partial^2 U}{\partial \phi^{*2}} &= \sum_{k=1}^T \binom{T}{k} P_k z^{k-3} (1 - z)^{T-k-2} \\ &\quad \times [(k - 1)(k - 2) - 2z(k - 1)(T - 2) \\ &\quad + z^2(T - 1)(T - 2)]. \end{aligned} \quad (44)$$

As stated above, if the  $\{P_k\}$  are parametrized in some way, and we are interested in finding at what parameter values two bifurcations appear and begin to separate, then we need to determine when  $\partial U / \partial \phi^* = 0$  [Eq. (35)] and when  $\partial^2 U / \partial \phi^{*2} = 0$  [Eq. (44)].

If, however, the  $\{P_k\}$  are fixed and we want to find all bifurcation points along with whether they are forward or backward bifurcations, then we can check the latter by finding the sign of  $d^2 p / d\phi^{*2} = 0$  [Eq. (43)]. Further analysis of these equations may be possible to find conditions on  $f$  and  $g$ , and thereby the  $\{P_k\}$ , that would ensure certain types of model behavior.

### Appendix B. Exact solution for $r < 1$ , $d^* = 2$ , and $T = 2$ and 3

For  $r < 1$ , we have individuals whose cumulative dose is below the threshold  $\bar{d}^*$  but are still infected because they have not yet recovered. To find the proportion of individuals in this category, we must calculate  $\Gamma_m$ , the fraction of individuals whose memory count  $D$  [number of successfully infecting interactions, Eq. (1)] last equaled the threshold  $m$  time steps ago and has been below the threshold since then. The fraction of these individuals still infected will be  $(1 - r)^m$ , i.e. those who have failed to recover at each subsequent time step. We write the proportion of infected individuals below the threshold as

$$\Gamma(p, \phi^*; r, T) = \sum_{m=1}^{\infty} (1 - r)^m \Gamma_m(p, \phi^*; T). \quad (45)$$

Once  $\Gamma$  is determined, a closed form expression for  $\phi^*(p)$  is obtained by inserting  $\Gamma$  into Eq. (15). To determine the  $\{\Gamma_m\}$ , we explicitly construct all allowable length  $m$  sequences of 1's and 0's such that no subsequence of length  $T$  has  $d^* = 2$  or more 1's. The analysis is similar for both the  $T = 2$  and 3 cases we consider in this appendix, and a generalization to all  $T$  is possible. Below, we first show the forms of  $\Gamma$  for the two cases and then provide details of the calculations involved.

For  $T = 3$ , we obtain

$$\begin{aligned} \Gamma(p, \phi^*; r, T) &= (p\phi^*)^2 (1 - p\phi^*)^2 \\ &\quad \times \left( 1 - r + \sum_{m=1}^{\infty} (1 - r)^m [\chi_{m-1}^{(3)} + \chi_{m-2}^{(3)} \right. \\ &\quad \left. + 2p\phi^* (1 - p\phi^*) \chi_{m-3}^{(3)} \right. \\ &\quad \left. + p\phi^* (1 - p\phi^*)^2 \chi_{m-4}^{(3)} \right], \end{aligned} \quad (46)$$

where  $\chi_m^{(T)}$  is defined as

$$\chi_m^{(T)}(p, \phi^*) = \sum_{k=0}^{\lfloor m/T \rfloor} \binom{m - (T - 1)k}{k} (1 - p\phi^*)^{m-k} (p\phi^*)^k. \quad (47)$$

Upon inserting  $\Gamma$  into Eq. (15), we have a closed form expression for  $\phi^*$  involving  $p$  and  $r$  as parameters. This expression can then be solved for numerically yielding the fixed point curves in Fig. 6.

For the  $T = \bar{d}^* = 2$  case, we find

$$\begin{aligned} \Gamma(p, \phi^*; r, T) &= (p\phi^*)^2 (1 - p\phi^*) \sum_{m=1}^{\infty} (1 - r)^m [\chi_{m-1}^{(2)} + p\phi^* \chi_{m-2}^{(2)}]. \end{aligned} \quad (48)$$

We consider the case of  $d^* = 2$  and  $T = 3$  first. For this specification of the model, there are two ways for an individual to transition to being below the threshold, i.e.

$D < d^*$ . An individual must have two positive signals and one null signal in their memory and then receive a null signal while losing a positive signal out the back of their memory window. The two sequences for which this happens are

$$\{d_{n-2}, d_{n-1}, d_n, d_{n+1}\} = \{1, 1, 0, 0\} \tag{49}$$

and

$$\{d_{n-2}, d_{n-1}, d_n, d_{n+1}\} = \{1, 0, 1, 0\} \tag{50}$$

with the point of transition to being below the threshold occurring between time steps  $n$  and  $n + 1$ . The two other sequences for which a node will be above the threshold are  $\{1, 1, 1\}$  and  $\{0, 1, 1\}$  but neither of these can drop below the threshold of  $d^* = 2$  in the next time step. Both the sequences of Eqs. (49) and (50) occur with probability  $(1 - p\phi^*)^2(p\phi^*)^2$ . When  $m > 1$ ,  $d_{n+2}$  may be either 0 or 1 for the first sequence, but for the second  $d_{n+2} = 0$  or otherwise the threshold will be reached again:

$$\{d_{n-2}, d_{n-1}, d_n, d_{n+1}, d_{n+2}\} = \{1, 0, 1, 0, 0\}. \tag{51}$$

Given these two possible starting points, we now calculate the number of paths for which  $D_{t+j}$  remains below  $d^* = 2$  for  $j = 1, \dots, m$ . The structure of an acceptable sequence must be such that whenever a 1 appears, it is followed by at least two 0's (otherwise,  $d^*$  will be exceeded). We can see therefore that every allowable sequence is constituted by only two distinct subsequences:  $a = \{0\}$  and  $b = \{1, 0, 0\}$ .

Our problem becomes one of counting how many ways there are to arrange a sequence of  $a$ 's and  $b$ 's given an overall sequence length  $m$  and that the length of  $a$  is 1 and the length of  $b$  is 3.

If we fix the number of subsequences of  $a$  and  $b$  at  $N_a$  and  $N_b$  then we must have  $m = N_a \cdot 1 + N_b \cdot 3$ . Varying  $N_b$  from 0 to  $[m/3]$  (the square brackets indicate the integer part is taken), we have  $N_a$  correspondingly varying from  $m$  down to  $m - 3[m/3]$  in steps of 3.

Next, we observe that the number of ways of arranging  $N_a + N_b$  subsequences  $a$  and  $b$  is

$$\binom{N_a + N_b}{N_a} = \binom{N_a + N_b}{N_b} = \frac{(N_a + N_b)!}{N_a!N_b!}. \tag{52}$$

To see this, consider a sequence of slots labeled 1 through  $N_a + N_b$ . In ordering the  $N_a$   $a$ 's and  $N_b$   $b$ 's, we are asking how many ways there are to choose  $N_a$  slots for the  $a$ 's (or equivalently  $N_b$  slots for the  $b$ 's). We are interested in the labels of the slots but not the order that we select them so we obtain the binomial coefficient of Eq. (52).

Allowing  $N_b$  and  $N_a$  to vary while holding  $m = N_a + 3N_b$  fixed, we find the total number of allowable

sequences to be

$$\sum_{N_b=0}^{[m/3]} \binom{N_b + N_a}{N_b} = \sum_{k=0}^{[m/3]} \binom{m - 2k}{k}, \tag{53}$$

where we have replaced  $N_b$  with  $k$  and  $N_a$  with  $3N_b - m = 3k - m$ . Noting that the probability of  $a$  is  $(1 - p\phi^*)$  and  $b$  is  $p\phi^*(1 - p\phi^*)^2$ , the total probability  $\chi_m^{(3)}(p, \phi^*)$  of all allowable sequences of length  $m$  for  $T = 3$  follows from Eq. (53):

$$\chi_m^{(3)}(p, \phi^*) = \sum_{k=0}^{[m/3]} \binom{m - 2k}{k} (1 - p\phi^*)^{m-k} (p\phi^*)^k. \tag{54}$$

For general  $T$ ,  $\chi_m^{(T)}$  is defined by Eq. (47).

We must also address some complications at the start and end of allowable sequences. At the end of a sequence, we have the issue of 1's being unable to appear because our component subsequences are  $a = \{0\}$  and  $b = \{1, 0, 0\}$ . The two endings we need to include are  $\{d_{n+m}\} = \{1\}$  and  $\{d_{n+m-1}, d_{n+m}\} = \{1, 0\}$ . We do this for each of the two starting sequences and we therefore have six possible constructions for allowable sequences of length  $m$ . For the starting sequence given in Eq. (49), we have the following three possibilities for  $\{d_{n-2}, d_{n-1}, d_n, d_{n+1}, \dots, d_{n+m}\}$ :

$$H_1 = \{1, 1, 0, 0, H_{m-1}^{a,b}\}, \tag{55}$$

$$H_2 = \{1, 1, 0, 0, H_{m-2}^{a,b}, 1\} \tag{56}$$

and

$$H_3 = \{1, 1, 0, 0, H_{m-3}^{a,b}, 1, 0\}, \tag{57}$$

where  $H_m^{a,b}$  is a length  $m$  sequence of  $a$ 's and  $b$ 's [which as we have deduced occur with probability  $\chi_m^{(3)}(p, \phi^*)$ ]. For the starting sequence given in Eq. (51), we similarly have

$$H_4 = \{1, 0, 1, 0, 0, H_{m-2}^{a,b}\}, \tag{58}$$

$$H_5 = \{1, 0, 1, 0, 0, H_{m-3}^{a,b}, 1\} \tag{59}$$

and

$$H_6 = \{1, 0, 1, 0, 0, H_{m-4}^{a,b}, 1, 0\}. \tag{60}$$

The probabilities corresponding to sequences (55) through (60) are

$$Pr(H_1) = (p\phi^*)^2(1 - p\phi^*)^2\chi_{m-1}^{(3)}(p, \phi^*), \tag{61}$$

$$Pr(H_2) = (p\phi^*)^3(1 - p\phi^*)^2\chi_{m-2}^{(3)}(p, \phi^*), \tag{62}$$

$$Pr(H_3) = (p\phi^*)^3(1 - p\phi^*)^3\chi_{m-3}^{(3)}(p, \phi^*), \tag{63}$$

$$Pr(H_4) = (p\phi^*)^2(1 - p\phi^*)^3\chi_{m-2}^{(3)}(p, \phi^*), \tag{64}$$

$$Pr(H_5) = (p\phi^*)^3(1 - p\phi^*)^3\chi_{m-3}^{(3)}(p, \phi^*) \tag{65}$$



and

$$Pr(H_6) = (p\phi^*)^3(1 - p\phi^*)^4\chi_{m-4}^{(3)}(p, \phi^*). \quad (66)$$

Summing these will give us the probability  $\Gamma_m$  but one small correction is needed for  $m = 1$ . By incorporating  $d_{n+2} = 0$  into the sequence of Eq. (51), we considered only  $m \geq 2$  sequences. So, we must also add in the probability of the  $m = 1$  sequence given in Eq. (50) which is  $(p\phi^*)^2(1 - p\phi^*)^2$ . Combining this additional quantity with the probabilities in Eqs. (61)–(66), we have

$$\begin{aligned} \Gamma_m(p, \phi^*, 3) &= (p\phi^*)^2(1 - p\phi^*)^2[\delta_{m1} + \chi_{m-1}^{(3)} + \chi_{m-2}^{(3)} \\ &\quad + 2p\phi^*(1 - p\phi^*)\chi_{m-3}^{(3)} + p\phi^*(1 - p\phi^*)^2\chi_{m-4}^{(3)}], \end{aligned} \quad (67)$$

where  $\delta_{ij}$  is the Kronecker delta function and we have suppressed the dependencies of the  $\chi_m^{(3)}$  on  $p$  and  $\phi^*$ . Finally, substituting this into Eq. (45), we obtain Eq. (46).

The calculation for  $T = 2$  follows along the same lines as above. We now have  $a = \{0\}$  and  $b = \{1, 0\}$  as our subsequences. There is only one starting sequence,

$$\{d_{n-1}, d_n, d_{n+1}\} = \{1, 1, 0\}, \quad (68)$$

and one ending sequences to add,  $\{d_{n+m}\} = \{1\}$  and  $\{d_{n+m}\} = \{1\}$ . Defining

$$\chi_m^{(2)} = \sum_{k=0}^{\lfloor m/2 \rfloor} \binom{m-k}{k} (1 - p\phi^*)^{m-k} (p\phi^*)^k, \quad (69)$$

the probability of being above the threshold and then having no reinfections may be written as

$$\Gamma_m(p, \phi^*; 2) = p^2\phi^{*2}(1 - p\phi^*)[\chi_{m-1}^{(2)} + p\phi^*\chi_{m-2}^{(2)}]. \quad (70)$$

Using Eq. (70) in Eq. (45) we obtain the form for  $\Gamma$  given in Eq. (48).

### Appendix C. Transition between classes I and II contagion models

The transition between classes I and II models of contagion occurs when a saddle-node bifurcation collides with the transcritical bifurcation lying on the  $p$ -axis. To find this transition, we must determine when the slope of the non-zero fixed point curve at the transcritical bifurcation (i.e. at  $p = p_c$  and  $\phi^* = 0$ ) becomes infinite. For the heterogeneous version of the model with  $r = 1$  and variable  $\{P_k\}$ , we are able to determine the behavior near  $p = p_c$  as follows. We first rearrange the right-hand side of Eq. (19) to obtain a

polynomial in  $p\phi^*$ :

$$\begin{aligned} \phi^* &= \sum_{k=1}^T \binom{T}{k} P_k (p\phi^*)^k (1 - p\phi^*)^{T-k} \\ &= \sum_{k=1}^T \binom{T}{k} P_k (p\phi^*)^k \sum_{j=0}^{T-k} \binom{T-k}{j} (-p\phi^*)^j \\ &= \sum_{k=1}^T \sum_{j=0}^{T-k} \binom{T}{k} \binom{T-k}{j} P_k (-1)^j (p\phi^*)^{k+j} \\ &= \sum_{l=1}^T \sum_{k=1}^m \binom{T}{k} \binom{T-k}{m-k} P_k (-1)^{m-k} (p\phi^*)^m \\ &= \sum_{m=1}^T C_m (p\phi^*)^m, \end{aligned} \quad (71)$$

where we have changed the summation over  $j$  and  $k$  to one over  $m = k + j$  and  $k$ . The coefficients  $C_m$  identified in the above may be written more simply as

$$C_m = (-1)^m \binom{T}{m} \sum_{k=1}^m (-1)^k \binom{m}{k} P_k, \quad (72)$$

since

$$\begin{aligned} \binom{T}{k} \binom{T-k}{m-k} &= \frac{T!}{k!(T-k)!} \frac{(T-k)!}{(m-k)!(T-l)!} \\ &= \frac{T!}{m!(T-m)!} \frac{m!}{k!(l-k)!} \\ &= \binom{T}{m} \binom{m}{k}. \end{aligned} \quad (73)$$

Expanding Eq. (71) to second order about  $p = p_c$  and  $\phi^* = 0$ , and writing  $\tilde{p} = p - p_c$ , we obtain

$$\phi^* \simeq C_1(\tilde{p} + p_c)\phi^* + C_2 p_c^2 \phi^{*2}. \quad (74)$$

From Eq. (72),  $C_1 = TP_1 (= 1/p_c)$  and  $C_2 = \binom{T}{2}(-2P_1 + P_2)$ . Using also that  $p_c = 1/(TP_1)$ , Eq. (74) then yields

$$\phi^* \simeq \frac{C_1}{C_2 p_c^2} \tilde{p} = \frac{T^2 P_1^3}{(T-1)(P_1 - P_2/2)} \tilde{p}. \quad (75)$$

Upon requiring  $d\phi^*/dp = \infty$ , the transition condition of Eq. (24) follows.

If  $C_2 = 0$  (i.e.  $P_1 = P_2/2$ ), the above calculation is no longer valid and the system exhibits a continuous phase transition with a non-trivial exponent at  $p_c$ . More generally, when  $C_2 = C_3 = \dots = C_n = 0$ , we find for small  $\phi^*$  and  $\tilde{p}$  that

$$\phi^* \simeq \frac{C_1}{C_{n+1} p_c^{n+1}} \tilde{p}^{1/n}. \quad (76)$$

We observe that for  $C_2 = C_3 = \dots = C_n = 0$ , a linearity condition in the  $\{P_k\}$  must hold, specifically,  $P_k = kP_1$  for  $k = 1, \dots, n$ . To show this, we rearrange the  $\{C_m\}$  given by Eq. (72) as follows, ignoring multiplicative

factors and substituting  $P_k = kP_1$ :

$$\begin{aligned} C_m &\propto \sum_{k=1}^m (-1)^k \binom{m}{k} P_k = \sum_{k=1}^m (-1)^k \frac{m!}{(k-1)!(m-k)!} P_1 \\ &\propto \sum_{k=1}^m (-1)^k \frac{(m-1)!}{(k-1)!(m-k)!} \propto \sum_{k'=0}^{m-1} (-1)^{k'} \binom{m-1}{k'} \\ &= [1 + (-1)]^{m-1} = 0, \end{aligned} \quad (77)$$

where we have shifted the index  $k$  to  $k' = k - 1$ .

## References

- Anderson, R.M., May, R.M., 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford.
- Banerjee, A.V., 1992. A simple model of herd behavior. *Quart. J. Econ.* 107, 797–817.
- Bass, F., 1969. A new product growth model for consumer durables. *Manage. Sci.* 15, 215–227.
- Bikhchandani, S., Hirshleifer, D., Welch, I., 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Polit. Econ.* 100, 992–1026.
- Brauer, F., Castillo-Chávez, C., 2001. *Mathematical Models in Population Biology and Epidemiology*. Springer, New York.
- Brock, W.A., Durlauf, S.N., 2001. Discrete choice with social interactions. *Rev. Econ. Stud.* 68, 235–260.
- Coleman, J., Menzel, H., Katz, E., 1966. *Medical Innovations: A Diffusion Study*. Bobbs Merrill, New York.
- Daley, D.J., Gani, J., 1999. *Epidemic Modelling: An Introduction*. Cambridge University Press, New York.
- Daley, D.J., Kendall, D.G., 1965. Stochastic rumours. *J. Inst. Math. Appl.* 1, 42–55.
- Diekmann, O., Heesterbeek, J.A.P., 2000. *Mathematical Epidemiology*. Wiley, New York.
- Dodds, P.S., Watts, D.J., 2004. Universal behavior in a generalized model of contagion. *Phys. Rev. Lett.* 92, 218701.
- Dushoff, J., Huang, W., Castillo-Chávez, C., 1998. Backwards bifurcations and catastrophe in simple models of fatal diseases. *J. Math. Biol.* 36, 227–248.
- Goffman, W., Newill, V.A., 1964. Generalization of epidemic theory: an application to the transmission of ideas. *Nature* 204, 225–228.
- Goldenfeld, N., 1992. Lectures on Phase Transitions and the Renormalization Group. *Frontiers in Physics*, vol. 85. Addison-Wesley, Reading, MA.
- Granovetter, M., 1978. Threshold models of collective behavior. *Am. J. Sociol.* 83, 1420–1443.
- Greenhalgh, D., Diekmann, O., de Jong, M.C.M., 2000. Subcritical endemic steady states in mathematical models for animals with infections with incomplete immunity. *Math. Biosci.* 165, 1–25.
- Haas, C.N., 2002. Conditional dose–response relationships for microorganisms: development and application. *Risk Anal.* 22, 455–463.
- Hadeler, K.P., Castillo-Chávez, C., 1995. A core group model for disease transmission. *Math. Biosci.* 128, 41–45.
- Hadeler, K.P., van den Driessche, P., 1997. Backward bifurcation in epidemic control. *Math. Biosci.* 146, 15–35.
- Hethcote, H., 2000. The mathematics of infectious diseases. *SIAM Rev.* 42, 599–653.
- Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. London. A* 115, 700–721.
- Kribs-Zaleta, C.M., 2002. Center manifold and normal forms in epidemic models. In: Castillo-Chávez, C., Blower, S., Kirschner, D., van den Driessche, P., Yakubu, A.-A. (Eds.), *Mathematical Approaches for Emerging and Re-emerging Infectious Diseases, An Introduction*, IMA Series in Mathematics and its Applications, vol. 126. Springer, New York, pp. 269–286.
- Kribs-Zaleta, C.M., Velasco-Hernandez, J.X., 2000. A simple vaccination model with multiple endemic states. *Math. Biosci.* 164, 183–201.
- Lohmann, S., 1994. The dynamics of informational cascades: the Monday demonstrations in Leipzig, East Germany, 1989–91. *World Polit.* 47, 42–101.
- Montgomery, R.L., 1996. *The Diffusion of Religions*. University Press of America, Lanham, MD.
- Morris, S., 2000. Contagion. *Rev. Econ. Stud.* 67, 57–78.
- Murray, J.D., 2002. *Mathematical Biology*, third ed. Springer, New York.
- Rogers, E., 1995. *The Diffusion of Innovations*, fifth ed. Free Press, New York.
- Schelling, T.C., 1973. Hockey helmets, concealed weapons, and daylight saving: a study of binary choices with externalities. *J. Conflict Resolution* 17, 381–428.
- Schelling, T.C., 1978. *Micromotives and Macrobehavior*. Norton, New York.
- Stark, R., 1996. Why religious movements succeed or fail: a revised general model. *J. Contemp. Religion* 11, 133–146.
- Strogatz, S.H., 1994. *Nonlinear Dynamics and Chaos*. Addison-Wesley, Reading, MA.
- Valente, T.W., 1995. *Network Models of the Diffusion of Innovations*. Hampton Press, Cresskill, NJ.
- Watts, D.J., 2002. A simple model of global cascades on random networks. *Proc. Natl Acad. Sci.* 99, 5766–5771.