OPEN

On the effect of heterogeneity in stochastic interacting-particle systems

Luis F. Lafuerza & Raul Toral

Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC), CSIC-UIB, Campus UIB, E-07122 Palma de Mallorca, Spain.

We study stochastic particle systems made up of heterogeneous units. We introduce a general framework suitable to analytically study this kind of systems and apply it to two particular models of interest in economy and epidemiology. We show that particle heterogeneity can enhance or decrease the size of the collective fluctuations depending on the system, and that it is possible to infer the degree and the form of the heterogeneity distribution in the system by measuring only global variables and their fluctuations. Our work shows that, in some cases, heterogeneity among the units composing a system can be fully taken into account without losing analytical tractability.

Most real systems are made up of heterogeneous units. Whether considering a population of cells[1,2], a group of people[3,4] or an array of lasers[5,6] (to name just a few examples), one never finds two units which behave exactly in the same way. Despite this general fact, quantitative modeling most often assumes identical units, since this condition seems necessary for having analytically tractable models. Moreover, in the general framework of complexity science, systems very often can be modeled only at a stochastic level, since a complete knowledge of all the variables, the precise dynamics of the units and the interaction with the environment is not available. One way to include system heterogeneity is to consider that the interactions between the units are not homogeneous but mediated by some complex network, an approach that has attracted enormous attention in the last years[7,8]. An issue that has been less studied, beyond the role of particle heterogeneity in deterministic systems[9–12], is the heterogeneity in the behavior of the particles themselves in stochastic models. Some exceptions include the recent reference[13], where the authors analyze the effect of heterogeneous transition rates on consensus times in the voter model, and works considering the effect of a few "committed" individuals in this and related models[14,15]. In the context of statistical physics, the combined effects of stochasticity and heterogeneity have been considered, for example, in random-field Ising models and spin glasses[16–18] or in diffusion in disordered media[19,20]. We aim here at developing a general framework for the analytical study of stochastic systems made up of heterogeneous units, applicable beyond equilibrium models or Hamiltonian systems and suitable for a general class of complex systems of recent interest, as well as at identifying some generic effects of particle heterogeneity on the macroscopic fluctuations.

In this work we will show that the combined effect of stochasticity and heterogeneity can give rise to unexpected, non-trivial, results. While, based on naïve arguments, one should conclude that global fluctuations increase in heterogeneous systems, we will show that in some systems of stochastic interacting particles fluctuations actually decrease with the degree of heterogeneity. Moreover, we will see that it is possible to infer the degree of particle heterogeneity (or "diversity") by measuring only global variables. This is an issue of great interest when one has access only to information at the macroscopic, population level, since it allows one to determine if heterogeneity is a relevant ingredient that needs to be included in the modeling. In this way, heterogeneity can be included when its presence is implied by the data and it does not enter as an extra free parameter. We will study first the simple case of independent particles; then we will consider the general case of interacting particles and develop an approximated method of general validity to analytically study these systems; next, as a way of example, this method will be applied to two particular models of interest in economy and epidemiology.

Our starting point is a stochastic description of a system composed by N non-identical units, which we call generically "particles" or "agents". Each particle is characterized by a constant parameter λ_i ($i = 1, \ldots, N$); the value of this parameter differs among the particles and it is the source of heterogeneity considered. Although there are more general ways of including heterogeneity, we will stick to this type of parametric heterogeneity[21] because it is simple yet rather general. For simplicity, we assume that each particle can be in one of two possible states and define $s_i(t) = 0, 1$ as the variable describing the state of particle i at time t (the two-states assumption will be relaxed later). The collective state of the system is given by the total number $n(t) = \sum_{i=1}^{N} s_i(t)$ of particles in

state 1. Sometimes, one does not have access to the individual dynamics and can only access experimentally the value of $n(t)$. We are interested in the statistical properties of this global variable and how do they depend on the degree of heterogeneity in the system. We will often refer to $n(t)$ as the *macroscopic* variable and to the $s_i(t)$'s as the *microscopic* ones.

Results

Independent particles. We study first the case in which particles jump independently from state 0 to 1 and vice-versa, schematically:

$$0 \xrightarrow{r_i^+} 1, \quad 1 \xrightarrow{r_i^-} 0 \qquad (1)$$

with rates that depend on the value of the heterogeneity parameter, $r_i^\pm = r^\pm(\lambda_i)$. The probability $p_i(t)$ for particle i to be in state 1 at time t obeys the linear rate equation $\frac{dp_i}{dt} = -r_i^- p_i + r_i^+ (1 - p_i)$. In the case of constant rates, the solution is: $p_i(t) = \frac{r_i^+}{r_i}(1 - e^{-r_i t}) + p_i(0)e^{-r_i t}$, with $r_i \equiv r_i^+ + r_i^-$. The results derived below apply equally if the rates depend on time or on the time that the particle has been in its current state (if the rate depends on the time a that the particle has been on its current state, the steady-state probability of finding the particle at state 1 is $p_{i,\text{st}} = \frac{\Lambda_i^-}{\Lambda_i^+ + \Lambda_i^-}$ with $\Lambda_i^\pm = \int_0^\infty dt\, e^{-\int_0^t da\, r_i^\pm(a)}$). Using particle independence and that the moments with respect to realizations of the stochastic process of the random variable s_i are given by $\langle s_i^k \rangle = 1^k p_i + 0^k(1 - p_i) = p_i$, one obtains that the average and variance of the global variable n are:

$$\langle n(t) \rangle = \sum_{i=1}^N p_i(t) = N\overline{p(t)}, \qquad (2)$$

$$\sigma^2[n(t)] = \sum_{i=1}^N \left(p_i(t) - p_i(t)^2\right) = N\left(\overline{p(t)} - \overline{p(t)^2}\right), \qquad (3)$$

where the overline denotes an average over the population, $\overline{g} \equiv \frac{1}{N}\sum g_i$. If we consider a system where all particles are identical (i.e. have the same values for the internal parameter $\lambda_i = \lambda_j, \forall i, j$), and keep the same average value $\langle n(t) \rangle$ for the global variable at time t, the variance would be $\sigma^2_{\text{id}}[n(t)] = N\overline{p(t)}\left(1 - \overline{p(t)}\right) \geq \sigma^2[n(t)]$. We obtain the somehow counterintuitive result that a system of heterogeneous independent particles displays smaller fluctuations in its collective variable than another system with identical particles. This effect is illustrated in figure 1. The reduction in the variance of the collective variable is N times the variance of p_i over the population:

$$\sigma^2_{\text{id}}[n(t)] - \sigma^2[n(t)] = N\left(\overline{p(t)^2} - \overline{p(t)}^2\right), \qquad (4)$$

which is of the same order, $O(N)$, as the variance itself, giving a non-negligible correction.

Reading the previous formula backwards, one realizes that the moments of the collective variable give information about the degree of heterogeneity in the system:

$$\overline{p(t)^2} - \overline{p(t)}^2 = \frac{\langle n(t) \rangle - \langle n(t) \rangle^2/N - \sigma^2[n(t)]}{N}. \qquad (5)$$

This expression is general, regardless the specific form in which p_i is distributed over the population. Higher moments of the heterogeneity distribution are also related to higher moments of the collective variable. This allows to infer the skewness, kurtosis and higher order characteristics of the heterogeneity distribution by measuring only global variables and their fluctuations. In the Supplementary Information it is shown that an equivalent result is obtained generically for M-state systems for $M > 2$.

Besides the moments, one can derive the full probability distribution of the global variable. The generating function for the one particle probability distribution is $g_i(z) = \sum_{s_i=0}^1 z^{s_i} P(s_i) = 1 - p_i + p_i z$ and the generating function $G(z) = \sum_{n=0}^N p(n)z^n$ of the sum of independent random variables is the product of their generating functions, $G(z) = \Pi_{i=1}^N g_i(z)$. Expanding in powers of z we can obtain the probability distribution for $n: P(n) = \frac{1}{n!(N-n)!}\sum_{i \in S_N} \Pi_{\ell=1}^n p_{i\ell} \Pi_{\ell=n+1}^N (1 - p_{i\ell})$, being $\mathbf{i} = (i_1, \ldots, i_N)$ and S_N the group of permutations of N elements.

The model studied in this subsection, despite its simplicity, offers a reduced description of generic systems of non-interacting multi-stable units subject to fluctuations. The results obtained here are directly relevant if one is interested in the collective properties of one such system when the units are non-identical. One can reasonably argue that the independence property is too unrealistic for the study to be of any practical interest. We will be considering more complicated cases including non-independent units in the rest of the paper. However, this simple model presents in isolation a mechanism, spontaneous transitions, that can play a role in more complicated and relevant systems (we will see this later). The simplicity of the model allows us to understand the effect of heterogeneity in this mechanism, and will give us insight in the role of heterogeneity in the behavior of more complicated systems.

Two types of uncertainties. We will now discuss the situation in which the particular values of the parameter of each particle are not
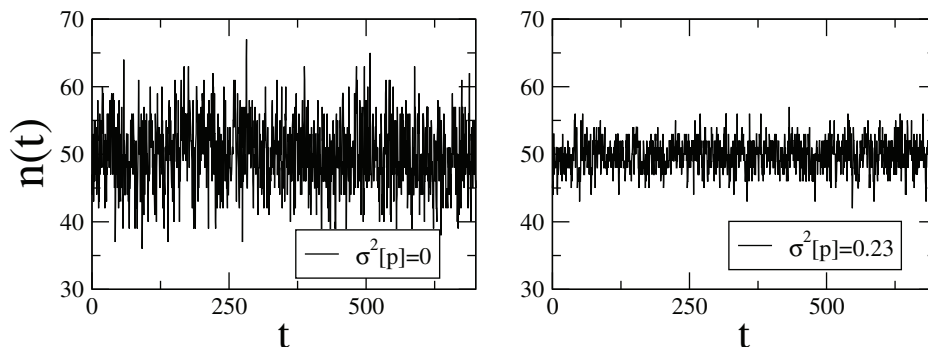
Figure 1 | Illustration of the effect of heterogeneity over the fluctuations. Time series for the global variable $n(t)$ of a system of identical (left panel) and heterogeneous (right panel) particles, for a system of $N = 100$ particles. The parameters were set as $r_i^+ = 1$, $r_i^- = 1/p_i - 1$, with $p_i = 1/2$ in the case of identical particles (left panel) and p_i chosen from a symmetric Beta distribution $f(p) = \frac{\Gamma(\alpha)^2}{\Gamma(2\alpha)}[p(1-p)]^{\alpha-1}$, with $\alpha = 0.05$, being the sample mean and variance equal to $\overline{p} = 0.501$, $\sigma^2[p] = 0.23$, respectively. Note that the fluctuations of the average state are larger in the case of identical particles.

known. This introduces an additional source of uncertainty. For simplicity, we will focus on the 2-states independent particles system considered before, but the discussion applies as well to general systems of interacting particles. This discussion will also allow us to take a closer look at the results obtained and clarify their meaning and relevance in different settings.

Often, one does not know the value of the parameter λ_i of each individual particle, but has some idea about how this parameter is distributed on the population, perhaps its probability distribution (obtained for example by measuring individual behavior in an equivalent system). Here, we will assume that the λ_i's are independent and identically distributed random variables with a given probability density $f(\lambda)$. In this case, $\langle n \rangle$ and $\sigma^2[n]$ are themselves random variables that, as shown above, depend on the particular realization of the λ_i's. The expected values of these quantities are obtained by averaging Eqs.(2,3) over the distribution of the individual parameters:

$$\langle \widehat{n(t)} \rangle = N\widehat{p(t)}, \quad \sigma^2\left[\widehat{n(t)}\right] = N\left(\widehat{p(t)} - \widehat{p(t)^2}\right), \qquad (6)$$

where the hat denotes an average with respect to f(λ), $\widehat{g} \equiv \int d\lambda g(\lambda) f(\lambda)$. Again the variance is smaller than for a system of identical particles with the same mean value, namely, $\sigma^2_{\mathrm{id}}[n(t)] - \sigma^2\left[\widehat{n(t)}\right] = N\left(\widehat{p(t)^2} - \widehat{p(t)}^2\right)$.

If we average the probability $P(n)$ over the distribution of parameters we obtain a simple form for the probability of the global variable n:

$$\widehat{P(n)} = \int d\lambda_1 \ldots d\lambda_N P(n|\lambda_1, \ldots, \lambda_N) f(\lambda_1) \ldots f(\lambda_N)$$
$$= \binom{N}{n}^n (1 - \hat{p})^{N-n}, \qquad (7)$$

a binomial distribution with parameter the average $\hat{p} = \int d\lambda p(\lambda) f(\lambda)$ over the distribution f. The variance of this distribution is

$$\sigma^2[n(t)]_{\mathrm{tot}} = N\left(\widehat{p(t)} - \widehat{p(t)}^2\right), \qquad (8)$$

equal to the variance one would obtain in a system of identical particles with the same average, $N\hat{p}(t)$, a result in apparent contradiction with (6). However, we should note that they refer to different things: Expression (6) gives the average variance when the parameter values are given, so measuring the average uncertainty in n due to the stochastic nature of the process. (8), in addition to the uncertainty coming from the stochasticity of the process, also includes the uncertainty on the parameter values.

The two expressions are related by the law of total variance:

$$\sigma^2[n]_{tot} = \widehat{\sigma^2[n]} + \sigma^2[\langle n|\lambda_1, \ldots, \lambda_N \rangle]$$
$$= N\left(\hat{p} - \widehat{p^2}\right) + \sigma^2\left[\sum_i p_i\right] = N\left(\hat{p} - \widehat{p^2}\right). \qquad (9)$$

In $\sigma^2[\langle n|\lambda_1, \ldots \lambda_N \rangle]$, the variances are taken over the distribution of the λ_i's. If we are considering a particular system, the temporal fluctuations (all the systems considered in this paper are ergodic, so we can think on averages over time or over the realization of the stochastic process interchangeably) in n will come only from the intrinsic stochasticity, and expressions (3,6) are the ones that measure it. Expressions (7,8) are appropriate only if we are considering an ensemble of systems with a distribution of parameters and our different measurements may come from different systems in the ensemble.

Formulation of the general method. Let us now consider a general system of interacting heterogeneous particles. The stochastic description now starts from a master equation for the N-particle probability distribution:

$$\frac{dP(s_1, \ldots, s_N)}{dt} = \sum_{j=1}^{N} (E_i - 1)\left[s_i r_i^- P(s_1, \ldots, s_N)\right]$$
$$+ \sum_{i=1}^{N} \left(E_i^{-1} - 1\right)\left[(1 - s_i) r_i^+ P(s_1, \ldots, s_N)\right], \qquad (10)$$

with step operators defined as $E_i^k F(s_1, \ldots, s_i, \ldots s_N) = F(s_1, \ldots, s_i + k, \ldots, s_N)$. The transition rates $r_i^\pm$ might now depend on the state of any other particle (this is how interactions enter in the model). From Eq.(10) one can derive for the moments and correlations:

$$\frac{d\langle s_i \rangle}{dt} = \langle r_i^+ \rangle - \langle (r_i^- + r_i^+) s_i \rangle \qquad (11)$$

$$\frac{d\langle s_i s_j \rangle}{dt} = -\langle q_{ij} s_j s_i \rangle + \langle r_i^+ s_j \rangle + \langle r_j^+ s_i \rangle \qquad (12)$$

with $q_{ij} = r_i^- + r_j^- + r_i^+ + r_j^+$ and $i \neq j$ in the second equation (recall that $s_i^2 = s_i$). In general, if the transition rates depend on the state variables s_i, these equations are not closed since they involve higher order moments, and some approximation method is needed to proceed. Systematic expansions in $1/N$, including van Kampen's Ω-expansion[22], are not applicable, since variables $s_i = 0, 1$ are not extensive. In the following, we introduce an approximation suitable for the analytical treatment of systems of globally coupled heterogeneous particles.

We assume that the m-particle correlations $\sigma_{j_1, \ldots, j_m}(t) = \langle \delta_{j_1}(t) \cdots \delta_{j_m}(t) \rangle$ with $\delta_j(t) = s_j(t) - \langle s_j(t) \rangle$ scale with system size as

$$\sigma_{j_1, \ldots, j_m}(t) = O\left(N^{-m/2}\right), \quad \text{for } j_k \neq j_l. \qquad (13)$$

Using this ansatz one can close the system of equations (11,12) for the mean values and the correlations. This is shown in the Supplementary Information for general transition rates of the form $f(s_1/N, \ldots, s_N/N)$.

While the resulting equations for the average values $\langle s_i(t) \rangle$ coincide with the mean-field rate equations usually formulated in a phenomenological way[12,17], our formulation allows us to compute the correlations and include, if needed, higher order corrections in a systematic way.

Assumption (13) can be justified noting that it is consistent with $\sum_{j_1, \ldots, j_m} \sigma_{j_1, \ldots, j_m} = \langle (n - \langle n \rangle)^m \rangle = O\left(N^{m/2}\right)$ which follows from van Kampen's splitting of the global variable $n = N\phi + N^{1/2}\xi$, with ϕ deterministic and ξ stochastic. Details are given in the Supplementary Information. The global variable n is extensive and it is expected to follow van Kampen's ansatz in many cases of interest. Note, however, that since there is not a closed description for the macroscopic variable n, one can not use van Kampen's expansion, and our approach extends the implications of this splitting of the macroscopic variable to the correlations of the microscopic state variables. For simplicity, we have focused on 2-states systems and assumed a constant number of particles. Systems with $M > 2$ states are also expected to follow ansatz (13), since the scaling of the global variable is not limited to 2-sates systems. The case of variable, but bounded, number of particles can be included straightforwardly by considering an extra state. The unbounded case can also be considered performing an appropriate limit. If the system has some spatial structure, the ansatz (13) is not expected to be valid, and some decay of the correlations with the distance is expected instead; the analysis of this interesting situation is left for future work.

We will proceed by applying the presented method to analyze the role of heterogeneity in two models previously considered in the literature that apply to contexts in which the assumption of identical agents can hardly be justified: stock markets and disease spreading. We will focus on the steady-state properties of both models, skipping transient dynamics.

Application to Kirman model. Kirman's model[23] was proposed to study herding behavior in the context of stock markets and collective dynamics on ant colonies. In the stock market context, agent i can be in two possible states (e.g. $0 \equiv$ "pessimistic" -with regard to future market price-and $1 \equiv$ "optimistic") and it can switch from one to the other through two mechanisms: spontaneous transitions at a rate ϵ, and induced transitions at a rate $N^{-1} \sum_j \lambda_j (1 - \delta_{s_i,s_j})$, being λ_j the "influence" of agent j on other agents. The case $\epsilon = 0$ corresponds to the voter model[24]. In the original formulation, all agents have the same influence, i.e. $\lambda_i = \lambda_j, \forall i, j$. We generalize the model allowing the parameter λ_i to vary between agents. In[25], the effect of heterogeneity was explored numerically, but not in a systematic way.

This model is interesting for us because it incorporates in a simple way two basic processes: spontaneous transitions and induced transitions. As we will see, due to its simplicity, a full analytical treatment is possible that will, in turn, allow us to obtain a deeper insight into the general effect of heterogeneity in systems of interacting particles.

The master equation for the process is of the form (10), with rates given by:

$$r_i^+ = \epsilon + N^{-1} \sum_k \lambda_k s_k, \quad r_i^- = \epsilon + N^{-1} \sum_k \lambda_k (1 - s_k) \quad (14)$$

From (11) the averages and correlations obey:

$$\frac{d\langle s_i \rangle}{dt} = \epsilon - (2\epsilon + \bar{\lambda})\langle s_i \rangle + N^{-1} \sum_k \lambda_k \langle s_k \rangle, \quad (15)$$

$$\frac{d\sigma_{i,j}}{dt} = -2(2\epsilon + \bar{\lambda})\sigma_{i,j} + N^{-1} \sum_k \lambda_k (\sigma_{i,k} + \sigma_{j,k}) \quad (16)$$

for $i \neq j$ and $\sigma_{i,i} = \langle s_i \rangle(1 - \langle s_i \rangle)$. Note that, due to the particular form of the rates, these equations do not involve higher-order moments. This is a simplifying feature of this model that allows one to obtain exact expressions. The first equation leads to a steady state value $\langle n \rangle_{st} = \frac{N}{2}$ (a property that comes from the symmetry $0 \leftrightarrow 1$). Using the relation $\sigma^2[n] = \sum_{i,j} \sigma_{i,j}$ we obtain (see Supplementary Information) that the variance in the steady state is:

$$\sigma_{st}^2[n] = \frac{N}{4}\left[1 + \frac{2\bar{\lambda}(1 - N^{-1})}{4\epsilon + \bar{\lambda}} + \frac{(N - 3 + 2N^{-1})\bar{A}}{2\epsilon + \bar{A}}\right] \quad (17)$$

with $A_i = \frac{\lambda_i^2}{N(4\epsilon + \bar{\lambda}) + 2\lambda_i}$. The leading-order term, $\sigma_{st}^2[n] = \frac{N}{4}\left[1 + \frac{\bar{\lambda}}{2\epsilon} + \frac{\sigma^2[\lambda]}{2\epsilon(4\epsilon + \bar{\lambda})}\right] + O(N^0)$, with $\sigma^2[\lambda] = \overline{\lambda^2} - \bar{\lambda}^2$, can also be readily obtained using the ansatz (13). Note that the presence of heterogeneity increases collective fluctuations. In figure 2 we compare expression (17) with results coming from numerical simulations.

In this case, the knowledge of $\langle n \rangle_{st}$ and $\sigma_{st}^2[n]$ alone does not allow to infer the degree of heterogeneity present in the system, unless one knows from other sources $\bar{\lambda}$ and ϵ and it is not possible to conclude whether the observed fluctuations have a contribution due to the heterogeneity of the agents. However, the steady-state correlation function $K[n](t) \equiv \langle n(t)n(0) \rangle_{st} - \langle n \rangle_{st}^2$, does include a term that allows to infer the possible heterogeneity. $K[n](t)$ is obtained integrating Eq.(15) and performing the appropriate conditional averages (see Supplementary Information):

$$K[n](t) = \left(\sigma_{st}^2[n] - u\right)e^{-(2\epsilon + \bar{\lambda})t} + ue^{-2\epsilon t}, \quad (18)$$

with $u \equiv \frac{2\epsilon + \bar{\lambda}}{\bar{\lambda}(1 - 1/N)}\left(\sigma_{st}^2[n] - N/4\right)$. The departure from a pure exponential decay signals the presence of heterogeneity (for identical particles $u = \sigma_{st}^2[n]$). Fitting this expression to data one can obtain $\sigma_{st}^2[n]$ and the parameters $\epsilon, \bar{\lambda}$. Then, the use of expression (17) would yield $\sigma^2[\lambda]$. In figure 3 we show that the numerical simulations indeed support the existence of two exponential decays for the correlation function.

Other ways to introduce heterogeneity. Interestingly, other ways to introduce heterogeneity in the system have different effects:

-If the heterogeneity is introduced in the spontaneous transition rate, $\epsilon \to \epsilon_i$, making some particles more prone to spontaneous transitions that others (but keeping $\lambda_j = \lambda, \forall j$ to isolate effects), collective fluctuations again increase with respect to the case of identical particles.
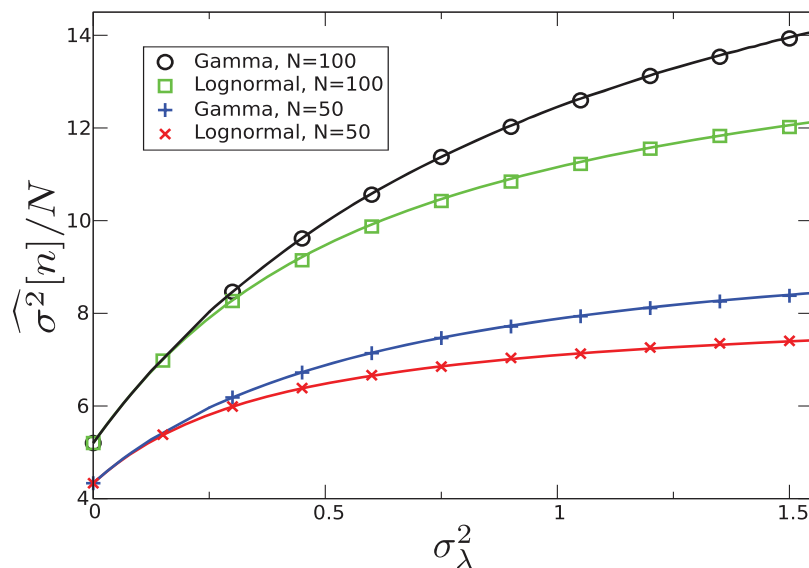
Figure 2 | Effect of the heterogeneity over the fluctuations in the Kirman model. Variance of the number of agents in state 1 as a function of the variance of the influence parameter λ in Kirman's model with distributed influence. Numerical simulations (symbols) and theoretical analysis (lines), Eq.(17), for different number of agents N and $\epsilon = 0.01$. λ_i are independent random variables distributed according to a lognormal or a gamma distribution with mean $\bar{\lambda} = 0.5$ and variance σ_λ^2. The results have been averaged over 2×10^4 for $N = 50$ and 10^4 for $N = 100$ realizations of the distribution of parameters.
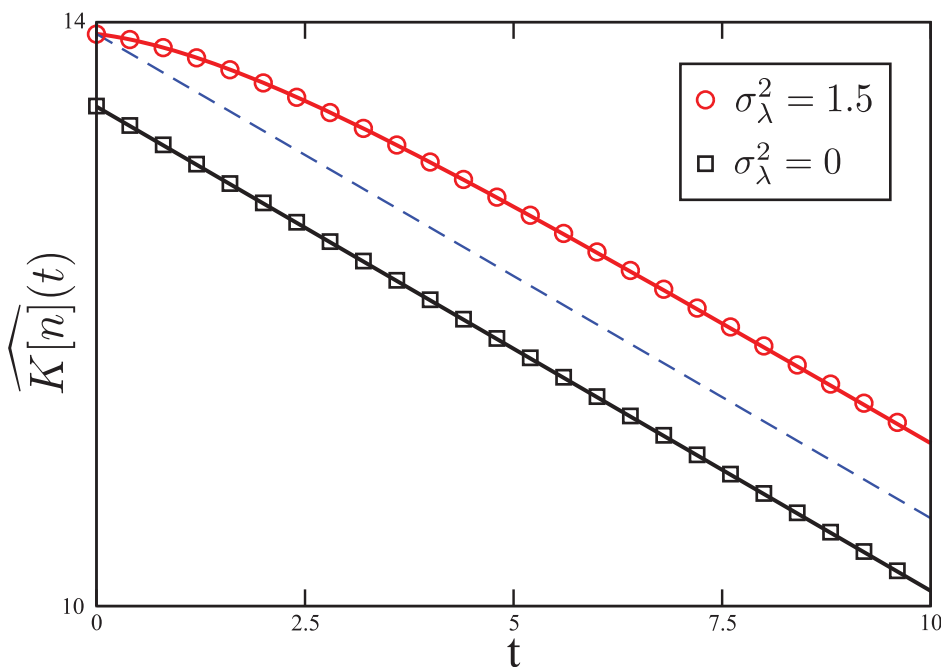
Figure 3 | Effect of the heterogeneity over the correlation function in the Kirman model. Correlation function (in log-linear scale) for Kirman's model with distributed influence. Results coming from numerical simulations (symbols) and theory (Eq.(18), solid lines). Note that when heterogeneity is present ($\sigma_\lambda^2 = 1.5$) the correlation function departs from purely exponential decay (displayed as a dashed line). Data for $\sigma_\lambda^2 = 0$ have been moved up 5.5 units vertically for better visualization. Parameters values are $\epsilon = 0.01$, $N = 100$. λ_i are independent random variable distributed according to a gamma with mean $\hat{\lambda} = 0.5$ and variance, σ_λ^2, indicated in the figure. A simple fit of expression (18) to the $\sigma_\lambda^2 = 1.5$ data gives $\overline{\lambda} = 0.50$, $\epsilon = 0.0099$.

-Next, we can assume that the rate of induced change is different for different agents, even if all have the same influence. Measuring this difference in "susceptibility" (to induced change) with a parameter ω_i, we would have that the rate of induced change in agent i is $\omega_i \sum_j \lambda_j (1 - \delta_{s_i, s_j}) / N$. The effect of heterogeneity in ω_i (keeping again $\lambda_j = \lambda$, $\forall j$) is that the collective fluctuations decrease with the degree of heterogeneity in the susceptibility ω_i.

-Setting some heterogeneous preference for the states among the particles, i.e. making ϵ_i^+, the spontaneous rate from 0 to 1 of particle i, different from ϵ_i^-, the spontaneous rate from 1 to 0 of the same particle, decreases global fluctuations. In order to vary the preference for one state keeping constant the global "intrinsic noise" of this particle (note that the correlation time of particle i, when isolated, is given by $\epsilon_i^+ + \epsilon_i^-$), we set $\epsilon_i^+ = 2\epsilon - \epsilon_i^-$ and generate ϵ_i^- as i.i.d. random variables with a distribution with support contained in the interval $[0, 2\epsilon]$. Exact explicit expressions for the first moments of the global variables are (see Supplementary Information):

$$\langle n \rangle_{st} = N \frac{\overline{\epsilon^+}}{2\epsilon} \qquad (19)$$

$$\sigma^2[n]_{st} = \frac{N}{4\left(\epsilon + \frac{2\lambda}{N}\right)} \left[\overline{\epsilon^+}\left(1 + \frac{\lambda}{\epsilon}\right) - \frac{\overline{\epsilon^+}^2}{\epsilon}\left(\frac{\lambda}{2\epsilon} + 1\right) - 2\frac{\sigma^2[\epsilon]}{2\epsilon + \lambda} \right], (20)$$

In figure 4 the exact expressions (19, 20) are compared with results coming from numerical simulations.

In this case, the correlation function decays exponentially,

$$K[n](t) = \sigma^2[n]_{st} e^{-2\epsilon t}, \qquad (21)$$

independently of the degree of heterogeneity, so this form of heterogeneity cannot be inferred by measuring the correlation function. Numerical simulations confirm this result.

Intuitive explanation of the effects of heterogeneity. We have seen that heterogeneity can have an ambivalent effect over the size of the

fluctuations, depending on the particular form it appears. We now provide intuitive arguments to understand these different effects.

When the influence parameter, λ_i, varies from one unit to the other, there will be some largely influential agents and others with little influence. In the limit of very large heterogeneity we can think of a situation with a single agent with an extremely large influence and the others having a negligible one (we are keeping a constant average influence). In this case, the highly influential agent drifts from one state to the other, essentially independently (since other agents have negligible influence), but, due to its large influence, all the agents are attracted to its current state. In this "follow the leader" regime, we obtain macroscopic transitions from one state to the other, corresponding to very large global fluctuations.

The situation is the opposite for a non-identical susceptibility parameter ω_i where global fluctuations decrease as the diversity is increased. Again, we can understand this in the limit of very large heterogeneity where a single agent (or a small number of them) has large susceptibility while all the others have a negligible one (in order to keep average susceptibility constant). Then, agents with small susceptibility change essentially independently, in an uncorrelated fashion, resulting in low global fluctuations (note that in order to have large global fluctuations, the fluctuations in the state of the single agents should be correlated).

In the case of diverse spontaneous transition rates, ϵ_i, global fluctuations increase with the degree of heterogeneity. In the limit of large heterogeneity, we would have a small number of agents with very large spontaneous transition rate, whose state would fluctuate in an uncorrelated fashion, and a large number of agents with low spontaneous transition rate, that essentially would only change state through induced transitions, giving rise to correlated fluctuations, resulting in large variance for the global variable.

In the case in which agents display an intrinsic heterogeneous preference for one of the two states, the global fluctuations decrease with heterogeneity degree. We saw this already in the first section for non-interacting agents. Here we see the same effect, suggesting that
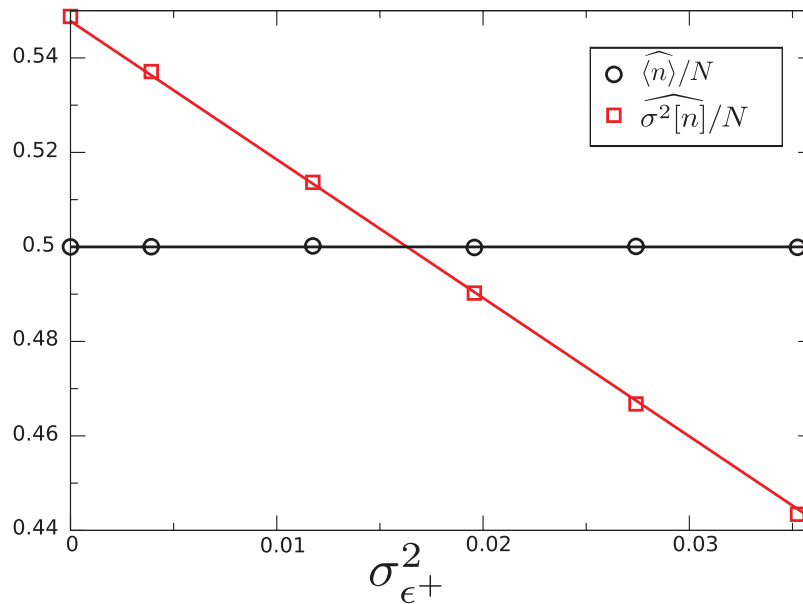
Figure 4 | Effect of the heterogeneity over the fluctuations in the Kirman model with distributed preference of states. Variance and average of the number of agents in state 1 as a function of the variance of the spontaneous transition rate to state 1, ϵ^+, in Kirman's model. Results coming from numerical simulations (symbols) and theoretical analysis (solid lines, Eqs.(19, 20)), for $N = 50$ agents, $\lambda_i = \lambda = 0.5$ and $\epsilon_+ + \epsilon_- = 2\epsilon = 0.4$. ϵ_i^+ are independent random variables distributed according to a symmetric beta distribution in the interval $(0, \epsilon)$ with mean $\overline{\epsilon^+} = 0.2$ and variance, $\sigma^2_{\epsilon^+}$.

the phenomenon is robust and still plays a role when interaction is added.

The coexistence of a small number of agents with large value for the parameter and a large number with a small value assumed in the previous arguments arises from the fact that an unbounded (from above) distribution of positive-defined parameters (e.g. rates) is skewed. However, all the effects of diversity commented are still present if the distribution is symmetric. In this case, nevertheless, the maximum degree of heterogeneity (for a constant mean value) is bounded, sometimes greatly limiting the maximum possible value of diversity. For symmetric distributions, a simple explanation is not so clear, but an asymmetry in the effect of increasing and decreasing the value of the parameter seems to be at the heart of the phenomenon.

Application to the SIS disease spreading model. The previous example could be treated exactly because, due to symmetry, the inter-action, non-linear terms, cancel out in the equations for the moments. In general, however, this is not the case, and the analytical treatment is more involved. Here we consider an example of such case. The stochastic susceptible-infected-susceptible (SIS) model is a paradigmatic model for the study of spreading of infectious disease[26] as well as the diffusion of innovation[11] and other types of social influence. Despite its simplicity, it captures interesting phenomen-ology. The process is schematically described by:

$$S(i)+I(j) \xrightarrow{\lambda_j/N} I(i)+I(j), \; I(j) \xrightarrow{\gamma} S(j), \; S(j) \xrightarrow{\epsilon} I(j), \; (22)$$
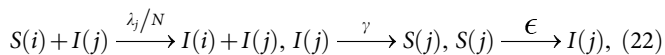
where S(i) (resp. I(i)) denotes agent i being susceptible (resp. infected). There are 3 basic elementary processes: (i) infected agent j infects susceptible agent i at a rate λ_j/N, being λ_j the infectivity parameter of agent j; (ii) infected agent j becomes susceptible a rate γ; (iii) susceptible agent j gets infected spontaneously (due to interactions with agents not considered in the system or other causes) at a rate ϵ. This corresponds to the SIS model with spontaneous contagions and distributed infectivity. In the absence of spontaneous infections $\epsilon = 0$, the system has a trivial steady state with zero infected agents. With $\epsilon \neq 0$ the system has a non-trivial steady state whose properties we analyze in the following. As in the previous case, heterogeneity could appear in any parameter of the

agents (for example, in the recovery rate, in a "susceptibility" parameter, etc.).

We study first the case in which only the infectivity, λ_i, can vary from agent to agent. The effect of heterogeneity in the deterministic version of related models was studied recently[12]. The master equation is of the form (10) with rates $r_i^+ = \epsilon + \sum_l \frac{\lambda_l \langle s_l \rangle}{N}$, $r_i^- = \gamma$. Equations (11–12) for the first moments can be closed in the steady state, using our main ansatz, to obtain explicit formulas for $\langle n \rangle_{st}$ and $\sigma^2[n]_{st}$ to any desired order in N^{-1}. In this case, however, the expressions are rather cumbersome and we skip them here. The results are plotted in figure 5, where we compare the approximation to order $O(N^{-1})$ with results coming from numerical simulations, showing good agreement. Here both the average value and the variance are modified by the presence of heterogeneity (the dependence of the average is, however, only in second order in $1/N$, almost unnoticeable in the figure). As in the Kirman model, the size of the fluctuations increases markedly with the amount of hetero-geneity in the "influence" (now influence to infecting others) of the agents.

In this case, other ways to introduce heterogeneity also have different effects. When heterogeneity appears in the recovery rate γ, the mean number of infected agent increases, with a moderate effect over the variance (resulting in smaller relative fluctuations). Heterogeneity in the susceptibility to infection (which would be introduced with the change $r_i^+ = \epsilon + \sum_l \frac{\lambda_l \langle s_l \rangle}{N} \to \epsilon + \omega_i \sum_l \frac{\lambda_l \langle s_l \rangle}{N}$, with ω_i distributed over the population) decreases the fluctuations, with little effect over the mean value. Heterogeneity in the spontan-eous infection rate ϵ has almost no effect. In a real situation, one expects to find heterogeneity simultaneously in several of the para-meters defining the model. When heterogeneity is present both in the infectivity and in the susceptibility, the effects of both types of het-erogeneity essentially add up, with the size of the fluctuations increasing with the heterogeneity in the infectivity for a given level of heterogeneity in the susceptibility and fluctuations decreasing with the level of heterogeneity in the susceptibility for a given level
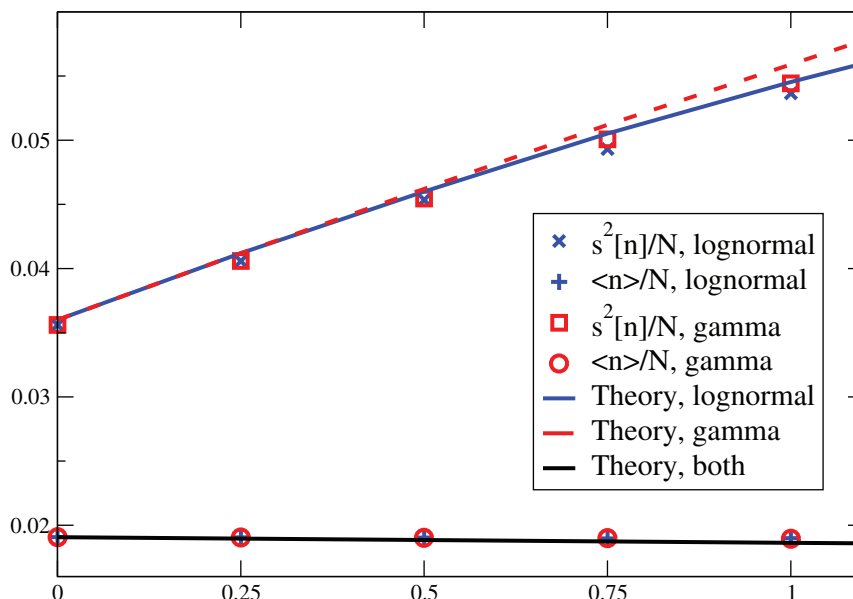
Figure 5 | Effect of the heterogeneity over the fluctuations in the SIS model. Average and variance of the number of infected agents in the SIS model as a function of the variance of the infectivity. Numerical simulations (symbols) and theoretical prediction to first order (lines). Parameters values are $\epsilon = 0.01$, $\gamma = 1$, $N = 200$, $\bar{\lambda} = 0.5$.

of heterogeneity in the infectivity. The effects of heterogeneity in the infectivity and in the susceptibility are equivalent to those found in the Kirman model, and can be intuitively understood in the same terms. Heterogeneity in the recovery rate is similar to assigning an heterogeneous preference for the state 0 (recovery) and its effect in the (relative) fluctuations is again the same as that in the case of the Kirman model. This suggests that the effects of the heterogeneity found are generic and can be useful to understand the behavior of other systems.

Discussion

In this work, we have analyzed the combined effect of stochasticity and heterogeneity in interacting-particle systems. We have presented a formulation of the problem in terms of master equations for the individual units, but extracted conclusions about the fluctuations of collective variables. We have developed an approximation suitable for the analytical study of this general type of systems. We have shown that the heterogeneity can have an ambivalent effect on the fluctuations, enhancing or decreasing them depending on the form of the system and the way heterogeneity is introduced. In the case of independent particles, heterogeneity in the parameters always decreases the size of the global fluctuations. We have also demonstrated that it is possible to obtain precise information about the degree and the form of the heterogeneity present in the system by measuring only global variables and their fluctuations, provided that the underlying dynamical equations are known. In this way stochastic modeling allows to obtain information not accessible from a purely deterministic approach. We have also demonstrated that, in some cases, one can account for the heterogeneity of the particles without losing analytical tractability.

Heterogeneity among the constituent units of a system is a very generic feature, present in many different contexts and this work provides a framework for the systematic study of the effect of heterogeneity in stochastic systems, having thus a wide range of potential applicability. More research in this direction would be welcomed.

Methods

We have developed and used analytical tools based on an extension of van Kampen's ansatz on the relative weight of the fluctuations compared to the mean value, suitable for systems with particle heterogeneity. We have included the details of this method in the Supplementary Information. In some cases, and in order to compare with the analytical expressions, we have generated data from numerical simulations using a particular form of Gillespie's algorithm[27] that takes into account the heterogeneity in the population. We now explain this algorithm using, for the sake of concreteness, the specific case of the Kirman model with distributed susceptibility.

The parameters of the system are: the spontaneous transition rate ϵ, the influence parameter λ, the susceptibility parameter of each agent ω_i, and the total number of agents N. In this case, the influence parameter can be reabsorbed rescaling ω_i, so we set $\lambda = 1$ without loss of generality. The variables of the system are the state of each agent $s_i = 0, 1$. We will also use the total number of agents in state 1, $n = \sum s_i$, the total susceptibility of agents in state 1, $\Omega = \sum s_i \omega_i$, and the average susceptibility $\bar{\omega} = \sum \frac{\omega_i}{N}$.

At any given instant, two events can happen:

(i) An agent in state 1 changes to state 0. This can happen due to a spontaneous transition, at a total rate $n\epsilon$, or due to an induced transition, at a total rate $\Omega \frac{(N-n)}{N}$.

(ii) An agent in state 0 changes to state 1. This can happen due to a spontaneous transition, at a total rate $(N-n)\epsilon$, or due to an induced transition, at a total rate $(N\bar{\omega} - \Omega) \frac{n}{N}$.

According to the Gillespie method, that considers the continuous-time process, the time at which the next transition will take place is exponentially distributed, with average the inverse of the total rate. The probability that a given transition is realized is proportional to its rate. If the realized transition is a spontaneous one, the agent that actually undergoes it is selected at random (since, in this case, they all have the same rate ϵ). If the transition is induced, the agent that undergoes it is selected with probability proportional to its susceptibility. It can be easily seen that this principles lead to an exact (up to numerical precision) simulation of sample paths of the stochastic process[27].

The algorithm, then, proceeds as follows:

(0) Evaluate the total number of particles in state 1, n, and the total susceptibility of particles in state 1, Ω.

(1) Evaluate the total transition rate $r = \epsilon N + \Omega \frac{N-n}{N} + (N\bar{\omega} - \Omega) \frac{n}{N}$.

(2) Generate the time for the next reaction, t_n, as an exponential random variable with average $1/r$. This can be done by setting $t_n = -\frac{1}{r} \ln U$, with U a uniform random variable in the range (0, 1).

(3) Select which reaction takes place. For this, generate a uniform random variable, g, in the range (0, r).

- If $g < n\epsilon$, the transition will be a spontaneous transition from 1 to 0; Select an agent, i, at random among those at state 1. Set $n = n - 1$, $\Omega = \Omega - \omega_i$.

- If $n\epsilon \leq g \leq N\epsilon$, the transition will be a spontaneous transition form 0 to 1; Select an agent, j, at random among those at state 0. Set $n = n + 1$, $\Omega = \Omega + \omega_j$.

- If $N\epsilon \leq g < N\epsilon + \dfrac{\Omega(N-n)}{N}$, the transition will be an induced transition from 1 to 0; Select an agent, j, among those at state 1 with probability proportional to the value of its susceptibility parameter ω_j. Set $n = n - 1$, $\Omega = \Omega - \omega_j$.

- If $g \geq N\epsilon + \dfrac{\Omega(N-n)}{N}$, the transition an induced transition from 0 to 1; Select an agent, j, among those at state 0 with probability proportional to the value of its susceptibility parameter ω_j. Set $n = n + 1$, $\Omega = \Omega + \omega_j$.

(4) Set $t = t + t_n$. Go to (1).

1. Spudich, J. L. & Koshland, Jr. D. E. Non-genetic individuality: chance in the single cell. *Nature* **262**, 467–471 (1976).
2. Snijder, B. & Pelkmans, L. Origins of regulated cell-to-cell variability. *Nat. Rev. Mol. Cell Biol* **12**, 119–125 (2011).
3. Granovetter, M. Threshold Models of Collective Behavior. *Am. J. Sociol.* **83**, 1420–1443 (1978).
4. Kirman, A. Whom or What Does the Representative Individual Represent? *J. Econ. Perspect* **6**, 117–136 (1992).
5. Braiman, Y., Kennedy, T. A. B., Wiesenfeld, K. & Khibnik, A. Entrainment of solid-state laser arrays. *Phys. Rev. A* **52**, 1500–1506 (1995).
6. Oliva, R. A. & Strogatz, S. H. Dynamics of a large array of globally coupled lasers with distributed frequencies. *Int. J. Bifurcat. Chaos* **11**, 2359–2374 (2001).
7. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
8. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. Complex networks: Structure and dynamics. *Phys. Rep.* **424**, 175–308 (2006).
9. Tessone, C. J., Mirasso, C. R., Tessone, C. J., Mirasso, C. R., Toral, R. & Gunton, J. D. Diversity-induced resonance. *Phys. Rev. Lett.* **97**, 194101 1–4 (2006).
10. Komin, N., Lacasa, L. & Toral, R. Critical behavior of a Ginzburg-Landau model with additive quenched noise. *J. Stat. Mech.* **P12008**, 1–19 (2010).
11. Peyton Young, H. Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning. *Am. Econ. Rev.* **99**, 1899–1924 (2009).
12. Novozhilov, A. S. Epidemiological Models With Parametric Heterogeneity: Deterministic Theory for Closed Populations. *Math Model Nat Pheno* **7**, 147–167 (2012).
13. Masuda, N., Gibert, N. & Redner, S. Heterogeneous voter models. *Phys. Rev. E* **82**, (010103R) 1–4 (2010).
14. Mobilia, M. & Georgiev, I. T. Voting and catalytic processes with inhomogeneities. *Phys. Rev. E* **71**, 046102 1–17 (2005).
15. Xie, J. & Sreenivasan, S. Social consensus through the influence of committed minorities. *Phys. Rev. E* **84**, 011130 1–8 (2011).
16. Fisher, D. S. Critical-behavior of random transverse-field Ising spin chains. *Phys. Rev. B* **51**, 6411–6461 (1995).
17. Young, A. P. (Ed.) *Spin Glasses and Random Fields* World Scientific, Singapore (1998).
18. Mézard, M., Parisi, G. & Virasoro, M. A. *Spin Glass Theory and Beyond* World Scientific, Singapore (1987).
19. Bouchaud, J.-P. & Georges, A. Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications. *Phys. Rep.* **195**, 127–293 (1990).
20. Ben-Avraham, D. & Havlin, S. *Diffusion and Reactions in Fractals and Disordered Systems* Cambridge University Press, Cambridge (2000).
21. Dushoff, J. Host Heterogeneity and Disease Endemicity: A Moment-Based Approach. *Theor Popul Biol* **56**, 325–335 (1999).
22. van Kampen, N. G. *Stochastic Processes in Physics and Chemistry*, North-Holland, Amsterdam (2004).
23. Kirman, A. Ants, Rationality, and Recruitment. *Q. J. Econ.* **108**, 137–156 (1993).
24. Liggett, T. M. *Interacting Particle Systems*, Springer-Verlag, New York (1985).
25. Alfarano, S. & Milaković, M. Network structure and N-dependence in agent-based herding models. *J. Econ. Dynam. Control* **33**, 78–92 (2009).
26. Anderson, R. M. *Population Dynamics of Infectious Diseases: Theory and Applications*, Chapman and Hall, London-New York (1982).
27. Gillespie, D. T. *Exact stochastic simulation of coupled chemical reactions* **81** (25), 2340–2361 (1977).

Acknowledgments

Author contributions

L.F.L. and R.T. designed the research, performed the research, contributed new analytic tools and wrote the paper.

Additional information