Stochastic evolution of bad memes

Ian Braga[*] and Lucas Wardil
Departamento de Física, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte MG, Brazil

(Received 24 March 2023; revised 11 July 2023; accepted 13 October 2023; published 4 December 2023)

Dawkins introduced a groundbreaking concept suggesting that humans, similar to other animals, operate as gene-propagating machines. Following in his footsteps, Blackmore posits that humans might distinguish themselves from other animals by also serving as specialized meme-replicating machines. Here we introduce a mathematical model that examines the impact of social conformity on the propagation of bad memes (memes with low intrinsic appeal). We state the *meme equations*, which give us the number of different kinds of memes living in the population and its total amount. We show that, unlike a virus, bad memes have a very low probability of initially spreading. However, as memes are produced in large numbers, some will eventually experience a stochastic rise and persist for extended periods, aided by social conformism within groups. We develop analytical approximations to calculate the mean time taken for memes to become extinct and the mean time spent in each population state. These approximations enable us to apply the meme equations to conduct a qualitative analysis.

DOI: 10.1103/PhysRevE.108.064103

I. INTRODUCTION

The term "meme" was introduced by Dawkins in the final section of his influential book, *The Selfish Gene* [1]. Dawkins proposes an analogous of the DNA information unit, the gene, but in the realm of ideas: "Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation."

The academic community met the meme concept with an initial skepticism [2]. However, nowadays, the internet provides numerous memes that can be monitored and analyzed for their distribution and fluctuations over time [3–7]. The meme concept is the base of a compelling hypothesis suggesting that the evolution of human intellect reached a turning point when humans acquired the ability to imitate others without discrimination, leading to the emergence of a new replicator that selects genes for producing brains with greater capacity to propagate memes [8]. Indeed, humans and primates process imitation through a neural mirroring system, simulating observed actions as if performed [9,10].

The study of meme propagation is often based on modifications of SIR models. Effective in diverse contexts, like internet search topics and Feynman diagram adoption in physics communities [11–14], these models overlook the distinction between virus and meme dissemination. Although there are "viral memes" like internet memes, newspaper headlines, popular songs, gossip, and emerging social truisms that quickly infect a large number of individuals and are quickly forgotten, there are also memes that are retained for longer times, like habits such as hand washing, shaving, kissing one's partner, congratulating someone on their birthday, wearing specific accessories, using cutlery during meals, nightly prayers, or playing chess. Moreover, the contagion mechanism is different. While virus spread as a cascade of infections, not all memes spread in this way. In [15], it was show that only a few very popular memes spread like infectious diseases. The majority of memes spread like complex contagions, where mechanisms like social reinforcement and homophily may be in place. So, the meme spreading mechanism is likely to be a combination of intrinsic quality and extrinsic mechanism, like the limited capacity of attention [16] or social effects [17].

There is another class of works that models the spread of ideas, but are not from the memes' point of view: the ideas, opinions, or cultural traits are fixed and the individuals adopt them through a social mechanism, like imitation. For example, in the famous Axelrod culture model [18], a predetermined number of features is proposed, each with a fixed number of traits. Consequently, memes are treated as static and almost inherent to human culture. These models are very similar to the well-established Ising spin model [19]: the two states correspond to the presence or not of the meme. The results are quite elegant: individuals exhibit random opinions at very high temperatures, several clusters of the same opinions are created at slightly above critical temperatures, yet no global magnetization is observed, and almost all individuals align in one direction below the critical temperature, indicating a ferromagnetic state. Random field Ising models have also been used, as in [20], where the authors predict a scaling law that is reasonably followed by three completely different kinds of social trend behaviors, namely, the evolution of the birth rates in European countries from 1960 to 2000, the adoption of cell phones in the latest 1990s, and the dynamics of an audience clapping.

Similar approaches have been employed in [21–23], which explain the diversity of opinions by considering spatial configurations or homophily as potential limiting factors. Additionally, variations of the branching process have been

[*] ianmbraga@gmail.com

utilized in other studies [24]. For instance, in [25], the authors analyze information cascades on Twitter and discover a reasonable alignment between the empirical data and the theoretical model. A comprehensive review of statistical mechanics applied to social dynamics is presented in [26].

Evolutionary game theory is also a suitable framework to analyze meme evolution if the goal is to investigate social factors [27]. This theoretical framework combines game-theoretical concepts, where the decisions of one agent are affected by the decisions of others, with stochastic death-birth evolution, which determines which types evolve in the population. Research often focuses on finding equilibrium concentrations of each type and on first-passage problems, such as determining the fixation probability and fixation time for each type. This framework is widely used to study the evolution of cooperation in population, but less used to analyze meme evolution. As one example of use, it is well known that the opinions of others, especially those in the same group, play a significant role in the adoption of memes, with individuals often attempting to conform to the majority view [28]. As a second example, in [29] the authors extend the Deffuant model by integrating game theoretical concepts of individual rational choices. However, there is not much done in the meme research field that uses evolutionary game theory as a theoretical tool.

Last, since interaction with others are important in the evolutionary game theory framework, the population structure is an important ingredient in the theory. The configuration in which the population is organized into groups, with each group forming a complete graph, has garnered notable attention in the existing literature [30–34]. An illustrative example of such a structured social environment can be found in schools, where individual classrooms constitute distinct groups. Within the same classroom, individuals often possess a stronger sense of self-identity and face similar social pressures, resulting in a higher likelihood of correlated behavior on average [35].

Here, we introduce a game-theoretical conceptualization of the meme invasion process to investigate the propagation of bad memes under social pressure in a population structured in groups. We define bad memes as memes that do not hold much appeal. For example, if the idea of self-harm is presented to a person in isolation, it would be regarded as a harmful meme since the individual would naturally have strong initial objections to it. However, when there is social pressure for conformity, which we model as a stag-hunt–like game [36], a meme has the potential to experience a stochastic rise, ultimately becoming dominant and persisting within a group. This can occur even if the meme lacks intrinsic appeal or quality. This phenomenon is of particular interest to the analysis of harmful memes because individuals are often subjected to the pluralistic ignorance effect, where harmful behavior is adopted on the grounds that others approve it, whereas the reality is the opposite: everyone is not comfortable with it [35].

More specifically, we posit that memes are continually generated by random individuals and, once introduced into the population, undergo distinct and parallel frequency-dependent Moran processes [37]. The birth rates of these memes are determined by their intrinsic quality and social pressure, which creates an invasion barrier. In this way, if many bad memes are produced, some may be subjected to a stochastic rise and overcome the initial barrier imposed by group conformism. Our model does not impose any restrictions on the number of memes that individuals can adopt, except for an intrinsic probability of being forgotten. Thus the only absorbent state is the one without the meme.

We calculate the mean extinction time for a meme with bad quality in populations of different sizes and social pressures. We also expand the formula to account for instances where the population is divided into groups of equal sizes, utilizing a time-scale separation technique. We also calculate the probabilities of a meme dominating i groups before becoming extinct and the conditional mean time that it remains in the population in such cases. Our approximations are validated through simulations, exhibiting a good level of agreement. Finally, we conduct a social analysis proposing the *meme equations* to quantify the amount of bad memes. The analysis confirms our hypothesis that the social factor can act as a reinforcement mechanism for unfavorable memes that occasionally flourish due to stochastic fluctuations.

The paper is organized as follows. In Sec. II, we introduce the stochastic evolutionary model. Section III analyzes the mean extinction time and the time that a single meme stays in each state. In Sec. IV, we make the social analysis of the model after defining the meme equations. Finally, Sec. V offers a general discussion of our findings.

II. MODEL

Let us suppose that a new meme, denoted by A, emerges in the mind of an individual within the population. An individual classified as type A carries and disseminates the meme A, while an individual classified as type B does not carry and does not propagate the meme A. It is important to note that type B does not represent an alternative meme, but rather the absence of an active meme A in one's mind. More specifically, a B individual either has never been in touch with the meme or has encountered it but refrains from propagating it.

We take a population of size N divided into n groups of equal size N/n. Initially, all individuals are of type B. Suppose a new meme A has emerged inside one individual's mind. Let N_{Ai} and N_{Bi} be, respectively, the number of A and B individuals in group i, for $i = 1, \dots, n$. We have n free variables because of the constraint $N_{Ai} + N_{Bi} = N/n$. Every time step one individual is randomly picked to change his idea. Let X be the type of this chosen individual and suppose that he pertains to group i. Then this individual imitates the strategy of a Y type individual of a group i with probability proportional to its fitness F_{Yi} and of a group $j \neq i$ with probability proportional to μF_{Yj}. The factor $\mu \in [0, 1]$ stands for the groups' connection. Also, if $X = A$, this individual has a probability proportional to γ to change to B without any imitation coming after. The parameter γ is the forgetting rate. A schematic illustration of the game-imitation dynamics is presented in Fig. 1.

The fitness of the memes is determined by their intrinsic quality a and by the payoff obtained in the interactions that the individuals carrying the meme have inside their groups. Supposing that the interactions in the groups are well mixed,

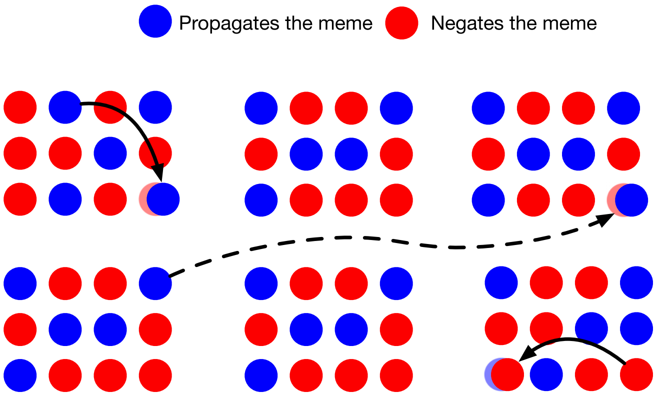

FIG. 1. Schematic illustration of the game and imitation dynamics. There are two types of individuals: those who have been exposed to the meme and actively disseminate it (blue) and those that negate the meme (red). In each time step, one individual is chosen to imitate another randomly chosen individual from his group (continuous arrow) or from another group (dashed arrow). The imitation probability depends on the frequency of each type (A and B) in the group: the greater the number of others of the same type, the higher the imitation probability of that type.

the fitness of Ai and Bi individuals are given by

$$F_{Ai} = a + sn\frac{N_{Ai}}{N},$$

$$F_{Bi} = 1 + s\left(1 - n\frac{N_{Ai}}{N}\right), \tag{1}$$

where a is the intrinsic quality of the meme and s is the intensity of the social factor. Note that the parameter a measures the intrinsic appeal of the meme, compared to its negation. For example, for $a = 0.1$ and $s = 0$, the individual has ten times more chance to imitate a B than an A individual. The memes that have an intrinsic value lower than that of their negation, that is, $a < 1$, are called *bad memes*. The social term in the fitness ranges from 0 to s as the fraction of A in the group ranges from 0 to 1. The game played inside each group is a stag hunt for $s > |1 - a|$, which is the appropriate game to model incentives of social conformity. Notice that the social factor is symmetric for A and B types. For $a = 1$, for example, the best strategy is the one that is the majority in the population, regardless of being A or B.

It is worth noting that certain models have the capacity to predict the distribution of various species through a straightforward dynamic process where types compete for space without requiring the assumption of varying fitness among them. For instance, Hubbell's Unified Neutral Theory of Biodiversity [38] aptly captures species distributions on islands. Moreover, the distribution of highly infectious memes' popularity appears to align well with models founded on the assumption of neutral fitness [39]. However, our focus in this study is directed towards highly retainable memes, typically adopted with more deliberation and within a framework of game-based interactions. This context entails that the likelihood of meme acceptance hinges on the group's state, representing an instance of *complex contagion* dynamics [15,17].

The state of the population is characterized by the vector $\mathbf{N} = (N_{A1}, \ldots, N_{An})$. The transition rates for our model are given by

$$T_i^{+}(\mathbf{N}) = \frac{1}{Z}\left(\frac{1}{n} - \frac{N_{Ai}}{N}\right)\sum_j \mu_{ij}\frac{N_{Aj}}{N}F_{Aj}, \tag{2}$$

$$T_i^{-}(\mathbf{N}) = \frac{1}{Z}\frac{N_{Ai}}{N}\left[\gamma + \sum_j \mu_{ij}\left(\frac{1}{n} - \frac{N_{Aj}}{N}\right)F_{Bj}\right], \tag{3}$$

where $\mu_{ii} = 1$ and $\mu_{ij} = \mu$ for $j \neq i$, and Z appears as a normalization factor obtained from $\sum_i (T_i^{+} + T_i^{-} + T_i^{0}) = 1$. The transition rate T_i^{0} corresponds to no transition.

Having a model, we can calculate the mean time of extinction of a meme with quality a given the fixed parameters N, n, s, γ, and μ. Notice that the social factor works simultaneously as a reinforcement for memes that successfully overcome the invasion barriers and as a resistance to the invasion of new ideas. In fact, if s and N/n are large enough, even memes with $a > 1$ have a low probability of initially spreading in the population, so the bad memes exist only due to the stochastic nature of the social dynamics.

Last, let us state clearly some assumption made in our model. First, we assume that the memes do not interact with one another in an individual's mind. As a result, each individual can carry multiple noninteracting memes and, for each meme, we classify individuals as either type A or type B. Second, we assume that all memes have the same intrinsic forgotten rate because we are concerned with the propagation of retainable memes. Moreover, we assume that γ is small and, since we would like to focus on the effects of social pressure and population structure, we assume that all memes have the same forgotten rate. Our mathematical analysis is based on the assumption of low connection between the groups. Nevertheless, we also perform simulations for high values of μ that help us to understand the range of validity of our approximation.

III. ANALYSIS OF SINGLE MEME INVASION

In this section, we begin by determining the average time it takes for a meme to become extinct when it originates from a single individual within a well-mixed population (a single large group). Our mathematical derivation for the scenario with a single group follows the methodology proposed in [40]. However, there is a difference to our model, as the state $N_A = 0$ is the only absorbing state. A comprehensive analysis of the one-dimensional case of the Moran process can be found in [41], where the author provides closed-form formulas for the mean fixation time and for the higher moments as well. Comprehensive resources on first passage problems and other frequently encountered inquiries concerning stochastic models can be found in [42,43].

In our study, there are n coupled random variables representing the fractions of memes in each group, which poses an analytical challenge. To extract meaningful insights from this complex system, we assume that the exchange of ideas between different groups is minimal. This enables us to employ a time-scale separation technique, allowing us to redefine the transition rates and conduct a mathematical analysis based on the single-group scenario. This approximation preserves the demographic effects of population fragmentation. We calculate the mean extinction time of the meme, the arrival probabilities at i groups, and the conditional time that a meme stays given that it has dominated i groups. While the equations apply to any population size, we illustrate the results only for small populations in order to compare with simulations, given that simulating fixation times in large populations is exceedingly time consuming.

A. Single group

First we analyze the dynamics when there is only one group. In this case our model is greatly simplified because we have only the variable N_A. The transition rates are given by

$$T^+(N_A) = \frac{1}{Z}\left(1 - \frac{N_A}{N}\right)\frac{N_A}{N}F_A, \tag{4}$$

$$T^-(N_A) = \frac{1}{Z}\frac{N_A}{N}\left[\gamma + \left(1 - \frac{N_A}{N}\right)F_B\right], \tag{5}$$

$$T^0(N_A) = \frac{1}{Z}\left[\left(\frac{N_A}{N}\right)^2 F_A + \left(1 - \frac{N_A}{N}\right)^2 F_B\right]. \tag{6}$$

Let $l = N_A$ and $T^*(N_A) = T_l^*$ for $* = +, -, 0$. Also, let t_l be the mean time of extinction of meme A when the population has l individuals of type A. It is true that $t_0 = 0$ and

$$t_l = 1 + T_l^+ t_{l+1} + T_l^- t_{l-1} + T_l^0 t_l \tag{7}$$

for $l = 1, \ldots, N$. Using that $T_l^0 = 1 - T_l^+ - T_l^-$, rearranging the terms, and defining $z_l = t_l - t_{l-1}$, we have

$$z_{l+1} = \lambda_l z_l - \frac{1}{T_l^+} \quad \text{with} \quad l = 1, \ldots, N, \tag{8}$$

where we defined $\lambda_l = T_l^-/T_l^+$. Because $t_0 = 0$, we have $z_1 = t_1$. The iteration of this equation yields

$$z_k = t_1 \prod_{p=1}^{k-1} \lambda_p - \sum_{p=1}^{k-1} \frac{1}{T_p^+} \prod_{q=p+1}^{k-1} \lambda_q, \tag{9}$$

for $k = 2, \ldots, N$. Summing z_k in Eq. (9) for k from $l + 1$ to N, and noticing that the sum applied to the definition of z_l is a telescopic one, we obtain

$$t_l = t_N - t_1 \sum_{k=l}^{N-1} \prod_{p=1}^{k} \lambda_p + \sum_{k=l}^{N-1} \sum_{p=1}^{k} \frac{1}{T_p^+} \prod_{q=p+1}^{k} \lambda_q. \tag{10}$$

This equation give us a relation between t_1 and t_N. We still need to look for another relation to link t_N with t_1. We take $l = N$ in Eq. (7) and rearrange the terms to obtain

$$t_N(1 - T_N^0) = 1 + T_N^- t_{N-1}, \tag{11}$$

where we recall that $T_N^+ = 0$, since the state $N + 1$ is not available. Finally, we just make $l = N - 1$ in Eq. (10), take the resulting expression of t_{N-1}, and plug it in Eq. (11). The variable t_N cancels out and we are left with the simplified equation

$$0 = 1 - T_N^- t_1 \prod_{p=1}^{N-1} \lambda_p + T_N^- \sum_{p=1}^{N-1} \frac{1}{T_p^+} \prod_{q=p+1}^{N-1} \lambda_q,$$

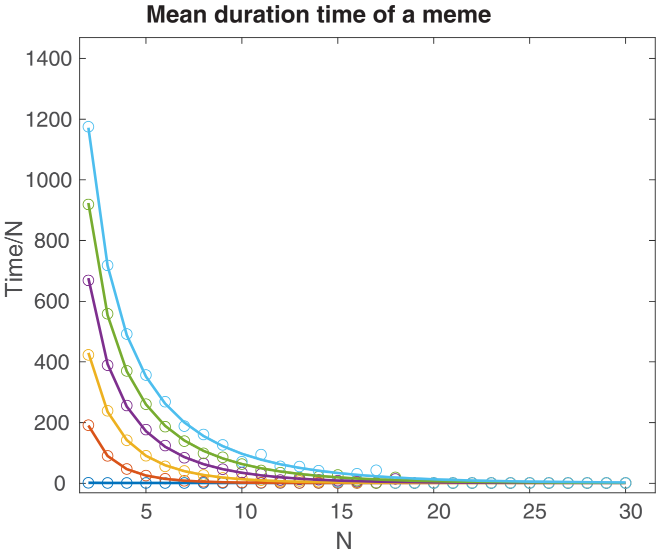

FIG. 2. Mean duration time of a bad meme. The y axis is the extinction time and the x axis is the group size. Parameters are $a = 0.1$ and $\gamma = 0.01$. The curves show the results for different s values: 0, 5, 10, 15, 20, and 25 (from bottom to top). Note that bad memes last for a much longer time in small groups and for high social terms. For example, for $s = 25$ the bad meme lasts almost 400 more times in a group of 5 people than it would last if $s = 0$ or if $N = 20$.

which gives

$$t_1 = \frac{\frac{1}{T_N^-} + \sum_{p=1}^{N-1} \frac{1}{T_p^+} \prod_{q=p+1}^{N-1} \lambda_q}{\prod_{p=1}^{N-1} \lambda_p}. \tag{12}$$

Equation (12) is the mean time it takes for a new meme to be extinct starting in one individual.

The expression for an arbitrary l is easy to find. Rearranging Eq. (10) for $l = 1$ we obtain t_N in the function of t_1:

$$t_N = t_1 + t_1 \sum_{k=1}^{N-1} \prod_{p=1}^{k} \lambda_p - \sum_{k=1}^{N-1} \sum_{p=1}^{k} \frac{1}{T_p^+} \prod_{q=p+1}^{k} \lambda_q.$$

Equation (10) provides a formula for t_l in the function of t_1 and t_N. Substituting the expression above in Eq. (10), and doing some algebra, we obtain

$$t_l = t_1 + t_1 \sum_{k=1}^{l-1} \prod_{p=1}^{k} \lambda_p - \sum_{k=1}^{l-1} \sum_{p=1}^{k} \frac{1}{T_p^+} \prod_{q=p+1}^{k} \lambda_q. \tag{13}$$

In Fig. 2, we compare the exact solution [Eq. (12)] to the Monte Carlo simulations for a meme with a small quality $a = 0.1$ for different social intensities. We re-scale the time to be inversely proportional to the group size, $\tau = t_1/N$. The reason is that, when the population grows, the rate at which the individuals are randomly picked to be updated has to grow proportionally, since the group size should not influence the rate at which the individuals change their ideas.

We draw attention to the significant effect that the social term has in favoring the entrance of bad memes. The average time a bad meme stands in the population can be two orders of magnitude greater for high values of s. Also, the effect depends on small populations, since they are more sensitive to stochastic fluctuations and the bad memes have a greater chance to invade by a sudden stochastic rise. Then, the social factor acts as a reinforcement, making the meme last long when it is by luck initially successfully spread.

B. Many groups

The analysis for many groups is more difficult because there are n coupled stochastic variables. However, in the limit of low group connections, $\mu << 1$, we can obtain an approximation for the mean extinction time of the meme. The social interpretation for this condition is that the groups are very strong unities. In such a scenario, individuals place significantly greater importance on the opinions of their fellow group members compared to the opinions of individuals outside their group.

We emphasize that there is a significant difference between the group-structured population with no connection, $\mu = 0$, and with very low connection, $\mu \ll 1$ [44]. If $\mu = 0$, there are n distinct absorbing states and n distinct quasistationary states. The states of each group after a relaxation time are independent. In contrast, if $\mu \ll 1$, there are only two states: the quasistationary state $N_A = N$ and the extinction state $N_A = 0$. In particular, we have shown that, even with very low connection, one meme can invade the entire population. This contrasts with the case where $\mu = 0$, where a single meme can invade only the group from which it originated.

In the regime $\mu << 1$, we can describe the system only in terms of the total number of A individuals, N_A. Because one individual is chosen at a time, the global transition rates are given by

$$T^*(\mathbf{N}) = \sum_i T_i^*(\mathbf{N}). \tag{14}$$

Suppose that we start the system with one A individual in some group that can be chosen to be the group 1. Suppose that after a time has passed, we have a quantity N_A of A individuals in the population. What is the most probable distribution of them in the groups?

Since the intergroup connection is very low, each group will evolve to its equilibrium much before the meme A can migrate to another group. Thus, if we have, for example, $N_A \leqslant N/n$ individuals A in the population, they will probably be in the group 1. Note that here we use the assumption that the forgotten rate is low $\gamma < 1/(N/n)$, which makes the state $N_{A1} = N/n$ a metaequilibrium for group 1. Suppose the population has $N/n < N_A \leqslant 2N/n$ individuals of type A. In this case, group 1 is probably complete, $N_{A1} = N/n$, and the remaining $N_A - N/n$ individuals are all located in one of the other groups. Because the global transition rates are symmetric by renumbering the group indices, we can assume without loss of generality that the remaining $N_A - N/n$ individuals A are all in the group 2. The same argument follows for all possible values of $1 \leqslant N_A \leqslant N$ and we can define the global transition rates for low group connection only in terms of N_A:

$$\mathcal{T}^*(N_A) = T^*(\mathbf{N}'), \tag{15}$$

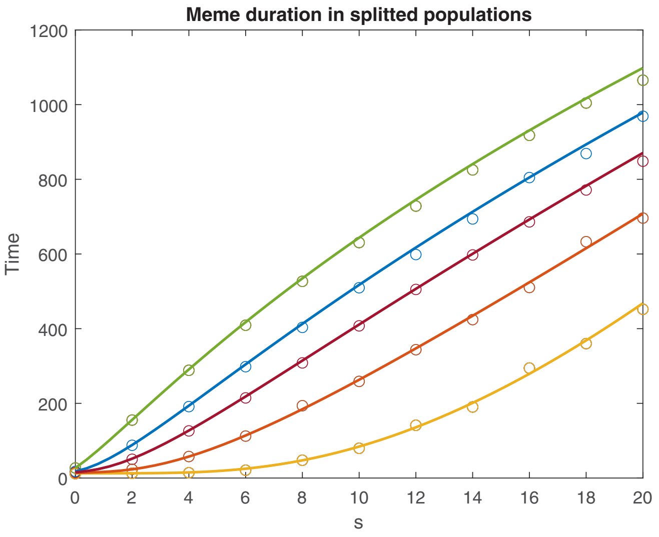

FIG. 3. Mean duration time of a bad meme in splitted populations. The y axis is the extinction time and the x axis is the social term s. Parameters are $N = 12$, $a = 0.1$, $\gamma = 0.01$, and $\mu = 0.001$. The curves show the results for different values of n: 1, 2, 3, 4, and 6 (from bottom to top). Note that bad memes last much longer in divided populations and for high social terms.

where $\mathbf{N}'$ is defined by

$$\mathbf{N}' = \begin{cases} (N_A, 0, \dots, 0) & \text{if } N_A \leqslant \frac{N}{n}, \\ \left(\frac{N}{n}, N_A - \frac{N}{n}, 0, \dots, 0\right) & \text{if } \frac{N}{n} < N_A \leqslant \frac{2N}{n}, \\ \vdots & \\ \left(\frac{N}{n}, \frac{N}{n}, \dots, N_A - \frac{(n-1)N}{n}\right) & \text{if } \frac{(n-1)N}{n} < N_A \leqslant N. \end{cases}$$

Finally, with these new rates defined, the analysis is analogous to the previous one done for a single group of size N. The mean time of extinction is given by the same equation [Eq. (12)], with the adjusted transition rates:

$$t_1 = \frac{\frac{1}{\mathcal{T}_N^-} + \sum_{p=1}^{N-1} \frac{1}{\mathcal{T}_p^+} \prod_{q=p+1}^{N-1} \Lambda_q}{\prod_{p=1}^{N-1} \Lambda_p}, \tag{16}$$

where $\Lambda_p = \mathcal{T}_p^- / \mathcal{T}_p^+$. Also, we have

$$t_l = t_1 + t_1 \sum_{k=1}^{l-1} \prod_{p=1}^{k} \Lambda_p - \sum_{k=1}^{l-1} \sum_{p=1}^{k} \frac{1}{\mathcal{T}_p^+} \prod_{q=p+1}^{k} \Lambda_q, \tag{17}$$

taking the new transition rates into Eq. (13).

In Fig. 3 we compare the approximation provided by Eq. (16) with simulations for a population of 12 individuals divided into 1, 2, 3, 4, or 6 weakly connected groups. We see that the social term enhances by far the mean extinction time of a very bad meme, that otherwise would rapidly be eliminated. Also, the effect is increased the more fragmented the population. The explanation, again, is due to the stochasticity intrinsic to small groups. To overcome the initial barrier of social conformity, the meme has to do a stochastic rise convincing up to 1/2 of the group. This improbable event becomes more probable if the groups are smaller so that the meme can enter bit by bit into the population. Once a bad

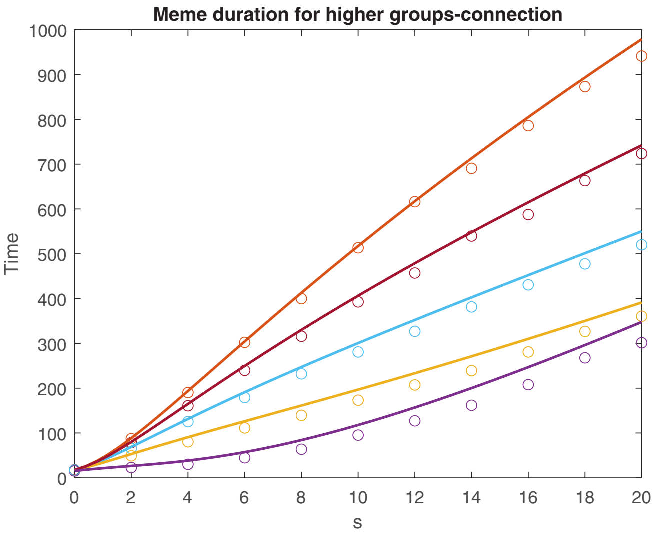

FIG. 4. Mean duration time of a bad meme in splitted populations for increasingly higher μ. The y axis is the extinction time and the x axis is the social term s. Parameters are $N = 12$, $a = 0.1$, $\gamma = 0.01$, and $n = 4$. From the top curve to the bottom the μ values are 0.001, 0.002, 0.004, 0.01, and 0.1. We can see that our analytical results depend on the condition of low migration, but they are surprisingly good approximations for the higher μ scenarios as well.

meme gets lucky, high social terms will ensure it a long duration.

Figure 4 also presents simulations of the mean extinction time for higher values of μ. Our analytical prediction starts to deviate for $\mu \geqslant 0.001$, but remarkably the approximation still captures well the mean extinction time for higher values of μ, far beyond the assumption that all groups are monomorphic when interchanging some meme. As μ increases, we observe that the mean extinction time of the meme becomes smaller. This suggests that, as the group-structured population becomes more connected, it resembles a scenario with only one group.

C. Arrival probabilities

In order to better understand the meme dynamics when transiting through the population, here we calculate the arrival probabilities of a meme that starts in one individual of the population, still in the regime of low migration where we use the single-variable transition rates $\mathcal{T}^*(N_A)$.

We call $P_i(l)$ the probability that the meme A invades i groups before being extinct, starting from l individuals. We can set the recursive relation

$$P_i(l) = \mathcal{T}^+(l)P_i(l+1) + \mathcal{T}^-(l)P_i(l-1) + \mathcal{T}^0(l)P_i(l), \tag{18}$$

with the boundary conditions $P_i(0) = 0$ and $P_i(iN/n) = 1$. Recalling that $\mathcal{T}^0(l) = 1 - \mathcal{T}^+(l) - \mathcal{T}^-(l)$, defining $z_i(l) = P_i(l+1) - P_i(l)$ and $\Lambda(l) = \mathcal{T}^-(l)/\mathcal{T}^+(l)$, we obtain

$$z_i(l) = \Lambda(l)z_i(l-1). \tag{19}$$

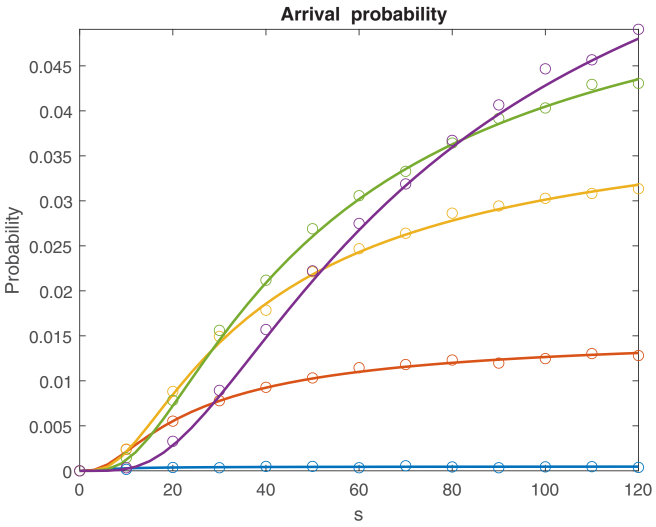

FIG. 5. Arrival probability of a bad meme. Here we plot the probability for the meme to invade a population of $N = 12$ individuals before being extinct. The y axis is the arrival probability and the x axis is the social term s. Parameters are $N = 12$, $a = 0.1$, $\gamma = 0.01$, and $\mu = 0.001$. The curves show the results for different values of n: 1, 2, 3, 4, and 6 (from bottom to top in the rightmost part of the graph). Note that bad memes have much more facility to invade when s is large and this effect is increased when the groups are more divided. However, there is an extensive range of s for which the invasion probability behaves nontrivially with the group division.

Summing in l from k to $iN/n - 1$ and recognizing that the sum over $z_i(l)$ is a telescopic one, we obtain

$$P_i(k) = 1 - \sum_{l=k}^{iN/n-1} z_i(l).$$

Calculating the sum on the right side using Eq. (19), we obtain the expression

$$P_i(k) = 1 - z_i(k)\left(1 + \sum_{j=k+1}^{iN/n-1} \prod_{l=k+1}^{j} \Lambda(l)\right).$$

Taking $k = 0$ we have $P_i(0) = 0$ and $z_i(0) = P_i(1) - P_i(0) = P_i(1)$, so that

$$P_i(1) = \frac{1}{1 + \sum_{j=1}^{iN/n-1} \prod_{l=1}^{j} \Lambda(l)}. \tag{20}$$

Finally, from Eq. (20) we can derive the probabilities of strict domination

$$P_i^{\text{only}} = P_i(1) - P_{i+1}(1), \tag{21}$$

which is the probability for the meme A to dominate i groups starting at one individual, but not dominate $i+1$ groups. Here we have that $i = 0, 1, \dots, n$, $P_{n+1}(1) = 0$, and $P_0^{\text{only}} = 1 - P_1(1)$ is the probability that no group is dominated by the meme A.

The comparison between Eq. (20) and simulations is provided in Fig. 5. We plot the probability of the meme invading a population of $N = 12$ individuals before being extinct for different group divisions. We can see that, although the groups

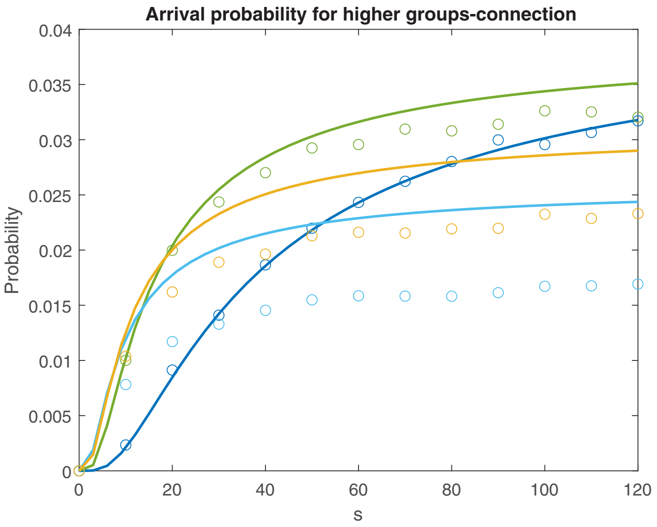

FIG. 6. Arrival probability of a bad meme for higher groups connection. Here we plot the probability for the meme to invade a population of $N = 12$ individuals before being extinct. The y axis is the arrival probability and the x axis is the social term s. Parameters are $N = 12$, $a = 0.1$, $\gamma = 0.01$, and $n = 3$. The curves show the results for different values of μ: 0.001 (blue), 0.01 (green), 0.05 (yellow), and 0.1 (light blue). Note that our analytical results start to deviate when the value of μ increases. Interestingly, we can see that there is a nontrivial value of μ for which the bad memes have more chance to invade in the presence of the social conformism.

are very weakly connected ($\mu = 0.001$), the fragmentation of the population may facilitate the total invasion of a very bad meme. We also show in Fig. 6 the arrival probability for higher values of group connection. We can see that our approximation strongly depends on the assumption of low connection. It is interesting to note that there is a counterbalance between two effects of the fragmentation. On one side it is more difficult for the meme to transit through the population. On the other side there is an advantageous effect by making the stochastic fluctuations more probable in the smaller groups, which helps the bad memes. Thus it is expected that the helpful influence of the fragmentation for the bad meme is increased when the social conformism is stronger, as shown in the figures. Also, we can see in Fig. 6 that there is a nontrivial value of μ for which the invasion probability is a maximum. This reveals the two opposite effects of population fragmentation in the propagation of a very bad meme.

D. Conditional times

We also calculate the average time it takes for the meme to be extinct starting from one individual, conditional to the fact that i groups were dominated, but not $i + 1$. This is called the conditional time of extinction t_i^{only}, with $i = 0, 1, \dots, n$. The meme's average duration when it is known that it has not dominated any group is t_0^{only}, in which case we assume that the meme remained all the time in only 1 individual.

To calculate the conditional times we set the equation

$$t(1) = P_0^{\text{only}} t_0^{\text{only}} + P_1^{\text{only}} t_1^{\text{only}} + \dots + P_n^{\text{only}} t_n^{\text{only}}, \tag{22}$$

where $t(1) = t_1$ is given in Eq. (16). Instead of trying to solve this equation, we can look at the following relation:

$$t(1) = P_0^{\text{only}} t_0^{\text{only}} + \left(1 - P_0^{\text{only}}\right) t_1(1), \tag{23}$$

where $t_i(1)$ is just the average time that the meme persists when it starts in 1 individual, conditional to the fact that i groups were dominated. Now let us estimate $t_1(1)$. First, we can write

$$t_1(1) = t_1^{\text{rlx}} + t(N/n), \tag{24}$$

where $t(N/n)$ is given by Eq. (17) and t_1^{rlx} is the relaxation time it takes for the meme A to dominate one group, knowing that the meme has achieved this goal. Note that the Markov propriety allows us to make this separation. Because we know that the meme invaded one group, we separate the time it takes for this invasion to happen, t_1^{rlx}, and then restart the system at the initial condition $N_A = N/n$. As we already have $t(l)$ for any l, we need to estimate t_1^{rlx}.

The assumption of bad memes and high social terms is important to make the analysis of t_1^{rlx} feasible. Any invasion of a bad meme, starting from one individual, is rare. When a rare event happens, like an extinction of a large population subjected to logistic transition rates [45], it will almost certainly take a deterministic path, which can be calculated by the approach of WKB theory for large N.

We provide a simple estimate of t_1^{rlx}. We assume that when it is known that the bad meme A dominated one or more groups, the path it took must not have any negative transition, where some A changed for B. Thus, in this path, we have $\mathcal{T}^{-}(l) = 0$. We then renormalize our rates and, for every state l of the population, the new transition rates are given by

$$\mathcal{T}_{\text{rlx}}^{+}(l) = \frac{\mathcal{T}^{+}(l)}{\mathcal{T}^{0}(l) + \mathcal{T}^{+}(l)}, \tag{25}$$

with $\mathcal{T}_{\text{rlx}}^{0}(l) = 1 - \mathcal{T}_{\text{rlx}}^{+}(l)$. There is an intuitive explanation for this approach. Since our memes have a very low probability of increasing in the population, principally in the initial state of a group invasion, the probabilities $\mathcal{T}^{+}(l)$ tend to be very small. Suppose that negative transitions $l \rightarrow l - 1$ happened in a path where an invasion took place. Then, the transition $\mathcal{T}^{+}(l - l)$ must have happened twice, multiplying one more very small probability to the path. Indeed, if more steps are taken, there are more ways to realize the invasion. However, this addition does not counterbalance the multiplication of more improbable steps in this case.

Then, when the population is in state l, it has for each time step the probability $\mathcal{T}_{\text{rlx}}^{+}(l)$ of transitioning for $l + 1$. This transition will take an average of $1/\mathcal{T}_{\text{rlx}}^{+}(l)$ time steps to happen. Thus, when the meme invades the population it stays approximately a time $1/\mathcal{T}_{\text{rlx}}^{+}(l)$ in l individuals. Finally, let us return to the calculation of t_1^{rlx}. It is easy to see that

$$t_1^{\text{rlx}} = \sum_{l=1}^{N/n-1} \frac{1}{\mathcal{T}_{\text{rlx}}^{+}(l)}, \tag{26}$$

which gives us $t_1(1)$, by Eq. (24), and enables us to obtain t_0^{only} by Eq. (23):

$$t_0^{\text{only}} = \frac{t(1) - \left(1 - P_0^{\text{only}}\right)t_1(1)}{P_0^{\text{only}}}. \tag{27}$$

Now we expand Eq. (23) one more term to write

$$t(1) = P_0^{\text{only}} t_0^{\text{only}} + P_1^{\text{only}} t_1^{\text{only}} + \left(1 - P_0^{\text{only}} - P_1^{\text{only}}\right)\left[t_2^{\text{rlx}} + t(2N/n)\right], \tag{28}$$

where

$$t_2^{\text{rlx}} = \sum_{l=1}^{2N/n-1} \frac{1}{\mathcal{T}_{\text{rlx}}^{+}(l)}. \tag{29}$$

The variable t_2^{rlx} is just the relaxation time it takes for the meme to dominate two groups, when it is known that it has done it. Since we already calculated t_0^{only}, our only unknown variable is t_1^{only}, which is given by

$$t_1^{\text{only}} = \frac{t(1) - P_0^{\text{only}} t_0^{\text{only}}}{P_1^{\text{only}}} - \frac{\left(1 - P_0^{\text{only}} - P_1^{\text{only}}\right)\left[t_2^{\text{rlx}} + t(2N/n)\right]}{P_1^{\text{only}}}. \tag{30}$$

In general, we have

$$t_i^{\text{only}} = \frac{t(1) - P_0^{\text{only}} t_0^{\text{only}} - \cdots - P_{i-1}^{\text{only}} t_{i-1}^{\text{only}}}{P_i^{\text{only}}} - \frac{\left(1 - P_0^{\text{only}} - \cdots - P_i^{\text{only}}\right)\left\{t_{i+1}^{\text{rlx}} + t[(i+1)N/n]\right\}}{P_i^{\text{only}}}, \tag{31}$$

for $i = 1, \dots, n-1$, and

$$t_n^{\text{only}} = \frac{t(1) - P_0^{\text{only}} t_0^{\text{only}} - \cdots - P_{n-1}^{\text{only}} t_{n-1}^{\text{only}}}{P_n^{\text{only}}}. \tag{32}$$

A comparison of this approximation for t_i^{only} with simulations is shown in Fig. 7. Although the approximation is not perfect, it fits reasonably well. More importantly, the approximation is sufficiently accurate to predict the effect of group fragmentation, allowing us to perform a qualitative analysis.

IV. MEME EQUATION AND SOCIAL ANALYSIS

In the previous section we analyzed the invasion of a single meme. Now let us analyze the scenario where multiple independent memes invade the population. Suppose that each individual of the population has the same density rate $r(a)$ of meme creation. Then, the infinitesimal rate of meme creation with quality between a and $a + da$ in a population of constant size N is given by

$$dR(a) = Nr(a)da. \tag{33}$$

Once a meme is created, initially in only one individual's mind, it can be rapidly eliminated or survive for a long time. However, in the limit of $t \to \infty$ it will certainly be eliminated.

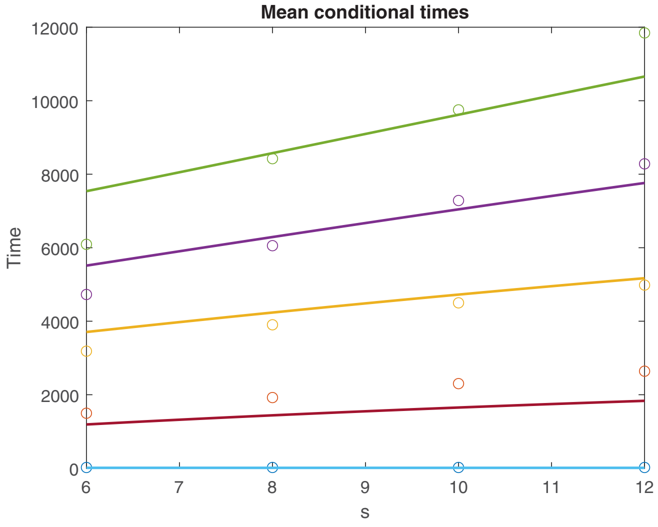

FIG. 7. Conditional extinction times. Here we plot the times for the meme to be extinct starting from one individual condition to the fact that it has dominated only 0, 1, 2, 3, and 4 groups, in a population of $N = 12$ divided into $n = 4$ groups. The y axis is the conditional time and the x axis is the social term s. Parameters are $N = 12$, $n = 4$, $a = 0.1$, $\gamma = 0.01$, and $\mu = 0.001$. From top to bottom, we show the conditional times of domination of only 4, 3, 2, 1, and 0 groups. The approximation gives us a great distinction of the orders of magnitude.

The reason is that memes have a forgotten rate γ. Thus the only attractor of the system is the state where the meme is extinct. We can then define a finite mean time to extinction of a meme with quality a, which we call $\tau(a)$. With these definitions, we can state the *meme equation*

$$dM(a) = \tau(a)dR(a), \tag{34}$$

which gives the infinitesimal amount of memes with quality between a and $a + da$ in the population. If we want to calculate, for example, the amount of bad memes in the population, $a \in [0, 1]$, we just have to integrate the expression in this interval to obtain

$$M_{[0,1]} = \int_0^1 dM(a) = \int_0^1 Nr(a)\tau(a)da. \tag{35}$$

This equation gives us a measure of the bad memes' variety, but we are also interested in the total amount of memes that are lying in all individuals minds, so we define the *total meme equation*, which gives the infinitesimal total amount of memes in the population:

$$dM_N(a) = \sum_{k=1}^{N} k\tau_k(a)dR(a), \tag{36}$$

where $\tau_k(a)$ is the average time that a meme with quality a stays occupying exactly k individuals. Clearly, we must have $\sum_k \tau_k(a) = \tau(a)$.

Taking our analytical derivations in the previous section, we have that $\tau = t_1$, given in Eq. (16). To state an approximation for the total meme equation, however, we need a little more elaboration. First we define the relaxation times of the forward and backward processes. The forward relaxation time is just the time it takes for the meme A to arrive at i groups, knowing that this happens:

$$t_i^{\mathrm{rlx}} = t_i^{\mathrm{go}} = \sum_{l=1}^{iN/n-1} \frac{1}{\mathcal{T}_{\mathrm{rlx}}^{+}(l)}.$$

The backward relaxation time is the time it takes for the meme A to go through its path of extinction starting with $iN/n - 1$ individuals knowing that it managed to dominate i groups:

$$t_i^{\mathrm{back}} = \sum_{l=1}^{iN/n-1} \frac{1}{\mathcal{T}_{\mathrm{rlx}}^{-}(l)}, \tag{37}$$

where $\mathcal{T}_{\mathrm{rlx}}^{-}(l)$ is defined in the same way as $\mathcal{T}_{\mathrm{rlx}}^{+}(l)$:

$$\mathcal{T}_{\mathrm{rlx}}^{-}(l) = \frac{\mathcal{T}^{-}(l)}{\mathcal{T}^{0}(l) + \mathcal{T}^{-}(l)}. \tag{38}$$

In other words, when i groups are dominated, we assume that the meme spent a time $1/\mathcal{T}_{\mathrm{rlx}}^{+}(l)$ in state l when it was going forward and a time $1/\mathcal{T}_{\mathrm{rlx}}^{-}(l)$ when it was going backward. Hence it takes a time $1/\mathcal{T}_{\mathrm{rlx}}^{-}(l) + 1/\mathcal{T}_{\mathrm{rlx}}^{-}(l)$ in the state l, for $l < iN/n - 1$, when it is known that i groups were dominated.

We can now state the approximation for the total meme equation [Eq. (36)]:

$$\begin{aligned}\frac{dM_N(a)}{dR(a)} &= P_0^{\mathrm{only}} t_0^{\mathrm{only}} \\ &\quad + \sum_{i=1}^{n} \sum_{l=1}^{iN/n-1} l P_i^{\mathrm{only}} \left(\frac{1}{\mathcal{T}_{\mathrm{rlx}}^{+}(l)} + \frac{1}{\mathcal{T}_{\mathrm{rlx}}^{-}(l)} \right) \\ &\quad + \sum_{i=1}^{n} \frac{iN}{n} P_i^{\mathrm{only}} \left(t_i^{\mathrm{only}} - t_i^{\mathrm{go}} - t_i^{\mathrm{back}} \right). \end{aligned} \tag{39}$$

At last, the density rate must be obtained experimentally; however, we can reasonably assume a natural functional dependence on a. Thus we can obtain the mean times for an evolving population following the Moran process and see how the group size and social pressure influence the abundance of bad memes for a given density rate distribution $r(a)$. We assume that the quality distribution of the density meme creation follows a half-normal distribution centered in $a = 0$ ($a \geqslant 0$), with variance σ^2:

$$r(a) = \frac{1}{\sigma} \exp\left(-\frac{1}{2} \frac{a^2}{\sigma^2} \right). \tag{40}$$

The variance σ^2 determines the society's quality in producing memes. In the limit of $\sigma^2 \to 0$, almost all memes produced are bad. In the limit $\sigma^2 \to \infty$, individuals produce memes with different qualities a at the same rate. Taking an intermediate value, $\sigma^2 = 10$, as our choice, we proceed by considering quenched disorder: we keep the creation rate fixed and study how the population structure and the social pressure influence the diffusion of bad memes.

First, we integrate the meme equation, Eq. (34), in the interval of all bad memes, $a \in [0, 1]$, for a population of $N = 12$ individuals. Figure 8 shows the numerical results. We note that, for this small population, the social term s increases the number of bad memes resident in the society. Also, we see that, for moderate s, there are more bad memes in split societies. However, for high values of s, the bad memes proliferate more in unified populations. The reason is that, although the probability of invasion becomes constant with increasing s, the conditional times become increasingly higher with s and the effect of bad memes duration in large groups is greater than the effect of groups splitting in the infiltration probabilities. Nevertheless, in Fig. 9, we integrate the meme equation for a population of size $N = 24$ and we can see that we need to have greater values of s for this nontrivial phenomenon to occur. Therefore, we can state that, when the populations are large, the population's fragmentation can only help the bad memes thrive for s that is not too large.

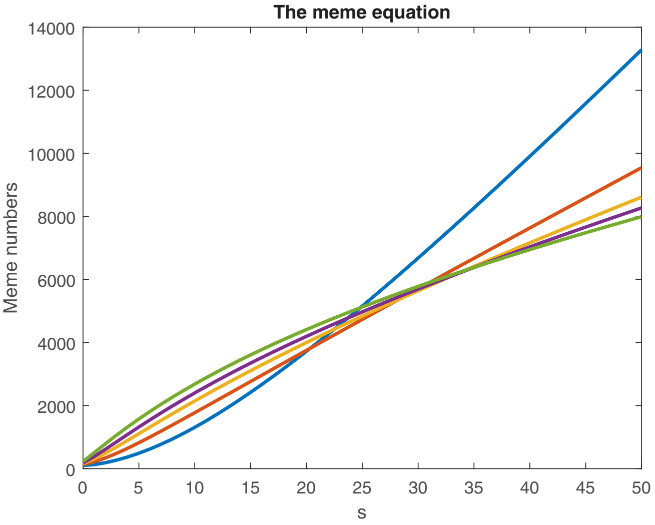

FIG. 8. Meme equation. Here we plot the meme equation over s, integrated in the interval of all bad memes $a \in [0, 1]$. Parameters are $N = 12$, $\gamma = 0.01$, and $\mu = 0.001$. The curves show the results for different values of n: 1, 2, 3, 4, and 6 (from bottom to top in the leftmost part of the graph). We see that the social term gives a great reinforcement for the bad memes and that split populations are more prone to them when s is moderate, but can act as well as a shield against bad memes when the social terms are disproportionately high.

Figure 10 shows the total meme equation plotted as a function of s for the same parameters of the plot in Fig. 8. This result enriches the analysis. First, as already expected, the increase of s increases the number of memes for all group divisions. Even more, although we saw that, for reasonable values of s, the division of the population increases the variety of bad memes, the total meme equation says that the total number of memes is higher for less divided populations. Note that the difference in the total number of memes is more pronounced for large s. Thus the fragmentation of the population diminishes the total number of bad memes, despite increasing their variety. When a meme manages to dominate one large group, it usually stays there a long time and for almost all this time it stays occupying the entire group.

We also show in Fig. 11 a plot of the meme equation for small s and large N. We can see an interesting phenomenon. For larger groups, the initial increase in s diminishes the number of bad memes. When N/n is larger, the increase in s makes it more difficult for the new memes to invade the

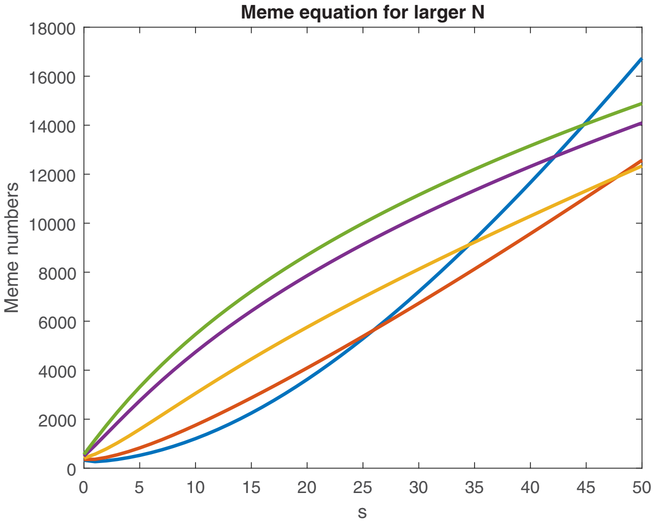

FIG. 9. Meme equation for larger N. Here we plot the meme equation over s, integrated in the interval of all bad memes $a \in [0, 1]$. Parameters are $N = 24$, $\gamma = 0.01$, and $\mu = 0.001$. The curves show the results for different values of n: 1, 2, 4, 8, and 12 (from bottom to top in the leftmost part of the graph). We see that, for a greater N, the split in the population favors the bad memes for reasonable values of s.

population because it always starts as a minority. However, from a certain value of s, the arrival probabilities no longer change with s. However, the conditional times keep growing until the increase of s returns to increase the number of bad memes.

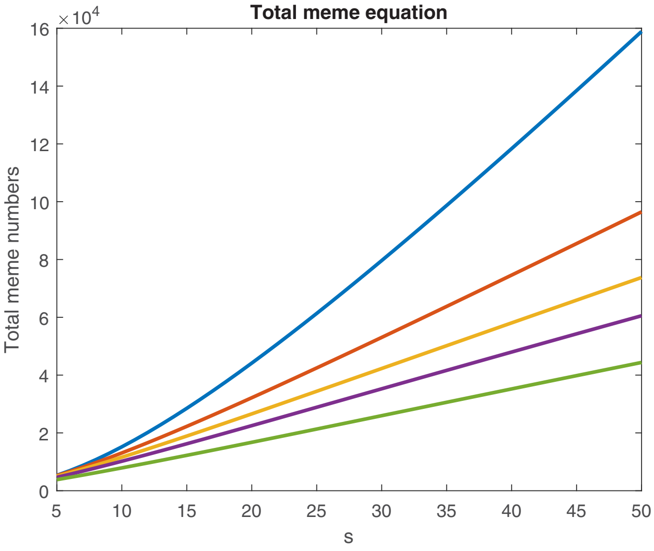

FIG. 10. Total meme equation. Here we plot the total meme equation over s, integrated in the interval of all bad memes $a \in [0, 1]$. Parameters are $N = 12$, $\gamma = 0.01$, and $\mu = 0.001$. The curves show the results for different values of n: 6, 4, 3, 2, and 1 (from bottom to top). Note that, for already moderate social terms, the total number of bad memes in the population greatly decreases with the population division.

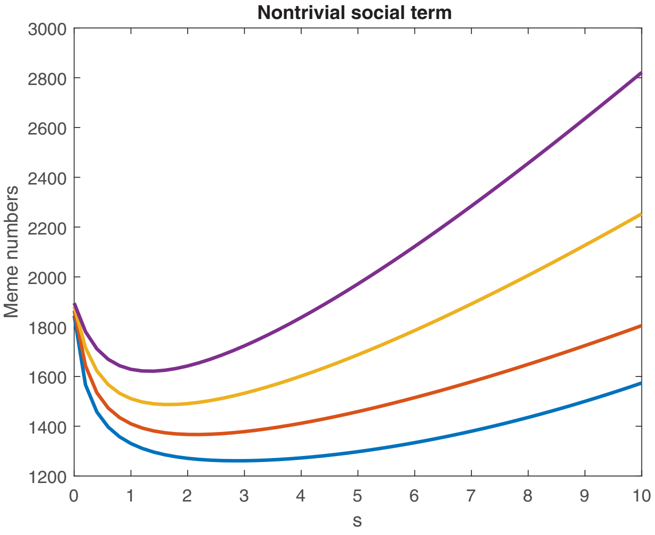

FIG. 11. Nontrivial social term. Here we plot the meme equation over s, integrated in the interval of all bad memes $a \in [0, 1]$. Parameters are $N = 60$, $\gamma = 0.01$, and $\mu = 0.001$. The curves show the results for different values of n: 1, 2, 3, and 4 (from bottom to top). At this scale, we see that there can be nontrivial social terms for which the bad memes' variety is a minimum.

V. CONCLUSIONS

The goal of this work is to understand how memes lacking inherent attractiveness can proliferate in a population due to stochastic occurrences of popularity and later be protected through conformity. We provide a mathematical analysis of the single meme invasion process and we introduce the *meme equations* to analyze the invasion of multiple memes. More specifically, we are interested in the effect of the size of the groups and the degree of the social pressure for conformism.

Using analytical techniques, we calculate the mean extinction time for a single meme that starts in one individual in a population composed of a single group. We also develop a good analytical approximation for scenarios involving multiple groups with limited intergroup connections. Lastly, we devise a more sophisticated approximation scheme to estimate the conditional times needed to integrate the total meme equation.

The assumption of independent evolution of memes employed here is also a simplification. Our model does not capture the phenomenon of memes correlation, described by Blackmore as *memeplexes* [8]. In reality, seemingly unrelated ideas can be linked in a critical way to drive major changes in behavior. For example, a study investigating political attitudes [27] demonstrated the existence of interconnected networks between different categories of attitudes towards a candidate. An individual's belief in a candidate's honesty, for instance, can influence their perception of the candidate's intelligence or popularity. Certain elements of attitudes were identified as more centrally connected, thereby yielding greater predictive power concerning individual voting behavior. Additionally, humans have a refined ability to sense social circles and gauge the prevailing state of opinion dynamics, thereby contributing to the correlation of memes. Once ideas become interlinked, they can serve as distinguishing markers of a well-defined social group [46]. Nonetheless, in this work, our goal is to understand the effects of group sizes and social pressure for conformity on the propagation of bad memes. Hence we assume that the memes evolve independent of each other. This simplification allows us to calculate the mean extinction time of an individual meme without considering the state of other meme dynamics. By employing this approach, we can focus on evaluating the influence of group size and social pressure on the propagation of individual memes in isolation.

The main empirical motivation for our investigation originates from a body of research spanning numerous decades, in which scholars have sought to grasp and numerically assess the impact of societal influence on the adoption of ideas. Some illustrative cases are provided by Latané's revision work [28]. In this revision, a series of experiments demonstrated substantial discrepancies in how individuals responded to the same questionnaire when answering alone versus in the presence of others. This work shows the substantial impact exerted by social conformity, which significantly shifts the probability distribution in favor of alternatives embraced by the majority. To capture these twofold effects, we introduce the parameters a and s.

The key contribution of this work lies in presenting a mathematical analysis that shows the substantial influence of social pressure for conformity on the propagation of memes that would otherwise face negligible chances of dissemination within the population. We elucidate how the intrinsic stochastic nature of social dynamics can underpin the emergence of unfavorable memes, particularly within more fragmented populations. These findings align harmoniously with empirical studies delving into the dynamics of less infectious memes [15,17]. These studies show that, in complex contagion scenarios, memes tend to propagate more effortlessly within disconnected populations, as they rely on social reinforcement for dissemination.

Our results bring attention to the significance of the memes that prevail in our society. Despite their potential misalignment with our best interests, their persistence can be attributed to stochastic events protected by conformism. For instance, even self-harm memes could find support within specific groups, despite their probable lack of appeal when considered in isolation [47]. As Blackmore points out, genes and memes do not consider the impact of their replication, as they cannot plan and make decisions based on the consequences of their actions. We cannot expect them to have created a desirable existence for us and, indeed, sometimes, they have not.

ACKNOWLEDGMENT

The authors thank the Brazilian agencies CNPq and CAPES, FAPEMIG-APQ-00694-18.

[1] R. Dawkins, *The Selfish Gene* (Oxford University Press, Oxford, UK, 1976).

[2] S. Blackmore, L. A. Dugatkin, R. Boyd, P. J. Richerson, and H. Plotkin, The power of memes, Sci. Am. **283**, 64 (2000).

[3] H. H. Chen, T. J. Tristram, D. F. M. Oliveira, and E. G. Altmann, Scaling laws and dynamics of hashtags on Twitter, Chaos **30**, 063112 (2020).

[4] L. A. Adamic, T. M. Lento, E. Adar, and C. Pauline, Information evolution in social networks, Proc. Ninth ACM Int. Conf. Web Search Data Mining **16**, 473 (2016).

[5] L. Weng, F. Menczer, and Y.-Y. Ahn, Predicting successful memes using network and community structure, *Proceedings of the International AAAI Conference on Web and Social Media* (2014), Vol. 8, pp. 535–544, https://ojs.aaai.org/index.php/ICWSM/article/view/14530.

[6] A. McNamara, Can we measure memes? Front. Evol. Neurosci. **3**, 1 (2011).

[7] C. M. Valensise *et al.*, Entropy and complexity unveil the landscape of memes evolution, Sci. Rep. **11**, 20022 (2021).

[8] S. Blackmore, *The Meme Machine* (Oxford University Press, Oxford, UK, 1999).

[9] G. Rizzolatti and M. A. Arbib, Language within our grasp, Trends Neurosci. **21**, 188 (1998).

[10] A. Whiten, D. M. Custance, J. C. Gomez, P. Teixidor, and K. A. Bard, Imitative learning of artificial fruit processing in children (Homo sapiens) and chimpanzees (Pan troglodytes), J. Comp. Psych. **110**, 3 (1996).

[11] R. Al-Amoudi, S. Al-Sheikh, and S. Al-Tuwairqi, Qualitative behavior of solutions to a mathematical model of memes transmission, Int. J. Appl. Math. Res. **3**, 36 (2014).

[12] R. Alghefari, S. Al-Sheikh, and S. Al-Tuwairqi, Global stability of stiflers impact on meme transmission model, Global J. Sci. Front. Res.: F Math. Decision Sci. **2014**, 9 (2014), https://www.researchgate.net/publication/303695635_Global_Stability_of_Stiflers_Impact_on_Meme_Transmission_Model.

[13] L. Wang and C. W. Brendan, An epidemiological approach to model the viral propagation of memes, Appl. Math. Modell. **35**, 5442 (2011).

[14] A. Cintrón-Arias, D. Kaiser, and C. Castillo-Chávez, The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models, Physica A **364**, 513 (2006).

[15] L. Weng, F. Menczer, and Y. Y. Ahn, Virality prediction and community structure in social networks, Sci. Rep. **3**, 2522 (2013).

[16] L. Weng *et al.*, Competition among memes in a world with limited attention, Sci. Rep. **2**, 335 (2012).

[17] D. Centola, The spread of behavior in an online social network experiment, Science **329**, 1194 (2010).

[18] R. Axelrod, The dissemination of culture: A model with local convergence and global polarization, J. Conflict Resolution **41**, 203 (1997).

[19] W. Weidlich, The statistical description of polarization phenomena in society, Br. J. Math. Stat. Psychol. **24**, 251 (1971).

[20] Q. Michard and J. P. Bouchaud, Theory of collective opinion shifts: From smooth trends to abrupt swings, Eur. Phys. J. B **47**, 151 (2005).

[21] M. Cinelli, M. G. De Francisci, A. Galeazzi, W. Quattrociocchi, and M. Starnini, The echo chamber effect on social media, Proc. Natl. Acad. Sci. USA **118**, e2023301118 (2021).

[22] A. Martins, Continuous opinions and discrete actions in opinion dynamics problems, Int. J. Mod. Phys. C **19**, 617 (2008).

[23] F. Bagnoli, T. Carletti, D. Fanelli, A. Guarino, and A. Guazzini, Dynamical affinity in opinion dynamics modeling, Phys. Rev. E **76**, 066105 (2007).

[24] R. Hofstad, *Random Graphs and Complex Networks* (Cambridge University Press, Cambridge, UK, 2016).

[25] J. P. Gleeson, T. Onaga, P. Fennell, J. Cotter, R. Burke, and D. J P O'Sullivan, Branching process descriptions of information cascades on Twitter, J. Complex Networks **8**, 1 (2021).

[26] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics, Rev. Mod. Phys. **81**, 591 (2009).

[27] J. Dalege *et al.*, Network structure explains the impact of attitudes on voting decisions, Sci. Rep. **7**, 4909 (2017).

[28] B. Latané, The psychology of social impact, Am. Psychol. **36**, 343 (1981).

[29] A. Di Mari and V. Latora, Opinion formation models based on game theory, Int. J. Mod. Phys. C **18**, 1377 (2007).

[30] I. Braga and L. Wardil, When stochasticity leads to cooperation, Phys. Rev. E **106**, 014112 (2022).

[31] R. M. May and M. Nowak, Superinfection, metapopulation dynamics, and the evolution of diversity, J. Theor. Biol. **170**, 95 (1994).

[32] C. Huia and M. A. McGeoch, Spatial patterns of prisoner's dilemma game in metapopulations, Bull. Math. Biol. **69**, 659 (2007).

[33] A. Eriksson and B. Mehlig, Metapopulation dynamics on the brink of extinction, Theor. Popul. Biol. **83**, 101 (2013).

[34] A. Traulsen, N. Shoresh, and M. Nowak, Analytical results for individual and group selection of any intensity, Bull. Math. Biol. **70**, 1410 (2008).

[35] C. M. Schroeder and D. A. Prentice, Exposing pluralistic ignorance to reduce alcohol use among college students, J. Appl. Social Psychol. **28**, 2150 (1998).

[36] M. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, Cambridge, MA, 2006).

[37] A. Traulsen, J. C. Claussen, and C. Hauert, Coevolutionary dynamics: From finite to infinite populations, Phys. Rev. Lett. **95**, 238701 (2005).

[38] I. Volkov *et al.*, Neutral theory and relative species abundance in ecology, Nature (London) **424**, 1035 (2003).

[39] Y. Kim, S. Park, and S.-H. Yook, The origin of the criticality in meme popularity distribution on complex networks, Sci. Rep. **6**, 23484 (2016).

[40] T. Antal and I. Scheuring, Fixation of strategies for an evolutionary game in finite populations, Bull. Math. Biol. **68**, 1923 (2006).

[41] P. Ashcroft, *The Statistical Physics of Fixation and Equilibration in Individual-Based Models*, Springer Theses (Springer, Cham, 2016).

[42] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 1992).

[43] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, UK, 2001).

[44] C. Hauert, Y. Chen, and L. Imhof, Fixation times in deme structured, finite populations with rare migration, J. Stat. Phys. **156**, 739 (2014).

[45] M. Assaf and B. Meerson, Extinction of metastable stochastic populations, Phys. Rev. E **81**, 021116 (2010).

[46] M. Galesic *et al.*, Human social sensing is an untapped resource for computational social science, Nature (London) **595**, 214 (2021).

[47] S. Sharma *et al.*, Detecting and understanding harmful memes: A survey, arXiv:2205.04274v2 [Int. Joint Conf. Artif. Intell. (to be published)].