Beyond six feet: The collective behavior of social distancing

Zhijun Wu

Department of Mathematics, Iowa State University, Ames, Iowa 50011, USA

*zhijun@iastate.edu

Abstract

In a severe epidemic such as the COVID-19 pandemic, social distancing can be a vital tool to stop the spread of the disease and save lives. However, social distancing may induce profound negative social/economic impacts as well. How to optimize social distancing is a serious social, political, as well as public health issue yet to be resolved. This work investigates social distancing with a focus on how every individual reacts to an epidemic, what role he/she plays in social distancing, and how every individual’s decision contributes to the action of the population and vice versa. Social distancing is thus modeled as a population game, where every individual makes decision on how to participate in a set of social activities, some with higher frequencies while others lower or completely avoided, to minimize his/her social contacts with least possible social/economic costs. An optimal distancing strategy is then obtained when the game reaches an equilibrium. The game is simulated with various realistic restraints including (i) when the population is distributed over a social network, and the decision of each individual is made through the interactions with his/her social neighbors; (ii) when the individuals in different social groups such as children vs. adults or the vaccinated vs. unprotected have different distancing preferences; (iii) when leadership plays a role in decision making, with a few leaders making decisions while the rest of the population just follow. The simulation results show how the distancing game is played out in each of these scenarios, reveal the conflicting yet cooperative nature of social distancing, and shed lights on a self-organizing, bottom-up perspective of distancing practices.

Introduction

In a severe epidemic such as the COVID-19 pandemic, social distancing can be a vital tool to stop the spread of the disease and save lives [1–13]. This is true before a vaccine becomes available and even after the population achieves certain herd immunity, for the pathogen always evolves, and the epidemic may develop into an endemic [14–22].

In general, social distancing is advocated as for people to keep certain distances from each other to reduce close social contacts [23–26]. However, in practice, it means a lot more and involves managing multiple social activities to avoid some of them but keep some others, in order to slow down the spread of infection while containing the negative social/economic impacts [27–51]. It is not clear how to optimize the activities though, often causing confusions and even rejections [52–58].

Some social activities have low contact rates but are socially isolating such as hiking, gardening, staying home, or reading. Some others are socially more involved
but have close social contacts such as large gatherings, going to night clubs, going to
shopping malls, or watching sports. In between, there are social activities that are
essential to our daily life such as grocery shopping, cafeteria dining, visiting friends,
taking buses, or going to schools or workplaces. Given a set of social activities, an
individual needs to make decision for how to participate in each of them, some
probably with higher frequencies while others lower or even completely avoided, so that
his/her close social contacts can be minimized at a least possible social/economic cost.

The decision of each individual must depend on or be influenced by the actions of
all other individuals in the population. For example, if everybody decides to stay
home, an individual may choose to watch a movie although the risk of having close
contacts at a movie theater is high. On the other hand, if the whole population
decides to go hiking, he/she may want to avoid it although hiking is usually a low
contact activity. Collectively, social distancing can thus be considered as a population
game \[59,60\], where based on what the population does, every individual makes
his/her own decision on how to participate in a given set of social activities so that
he/she can minimize his/her social contacts and possible social/economic costs. An
optimal distancing strategy can then be obtained when the game reaches an
equilibrium.

Research on social distancing has surged since the outbreak of the COVID-19
pandemic, providing a wealth of knowledge and experience on non-pharmaceutical
measures for preventing epidemics from spreading. However, most of these studies are
about the influences of public policies, economic concerns, or cultural differences on
social distancing, but not particularly about social distancing as a human behavior or
the collective behavior of a population for that matter. Work on modeling social
distancing has been pursued in the past \[61–64\] including some done recently \[65–78\],
but the focus is on the dynamics of epidemics with changing patterns or levels of
social distancing, with little specifics on how the distancing activities are carried out
and how certain distancing patterns or levels are achieved.

The work in this paper follows a game theoretic approach to social behavior in
general \[79–84\] and to social distancing in particular \[85,88\], and investigates social
distancing with a focus on how every individual reacts to an epidemic, what role
he/she plays in social distancing, and how the individual decision contributes to the
action of the population and vice versa. The collective behavior of social distancing is
modeled as a population game, where every individual makes a distancing decision
and together the population reaches an equilibrium when every individual achieves
his/her distancing goal.

A number of issues rise immediately for the general game model for social
distancing: First, a general population game assumes that every individual interacts
with all others in the population and knows their strategies, which is not true in the
real world, where people usually interact only with their social acquaintances. The
game also assumes that the individuals are all the same when evaluating the
distancing risks and making their distancing decisions, but in reality, they are not. For
example, children and adults seem to have different infection rates for COVID-19 and
would therefore consider the distancing risks of social activities differently; and so do
the vaccinated and unprotected individuals. The general game also requires every
individual to be able to make rational decisions for the game to eventually reach
equilibrium. The condition is again unrealistic, for not everyone is able to or willing to
make his/her own decisions.

However, all these issues can be addressed by making several refinements on the
general model: First, the population can be assumed to be distributed over a social
network, and the decision of each individual can be made through the interactions
with his/her social neighbors. Such a network can be simulated by generating a
small-world network using for example the Watts-Strogatz algorithm [89]. The simulation results presented in this paper show that the distancing game can be played successfully on such a network. Surprisingly, the game approaches to an equilibrium state as in the general case even when the interactions among the individuals are restricted only to their close neighbors.

Second, the population can be divided into different social groups according to certain social/biological/medical characteristics such as the age of the individuals (e.g., children, adults, seniors, etc.) or the level of protection (e.g., the vaccinated, recovered, unprotected, etc.) or the economic vulnerability (income above average, middle income, low income, etc.). The distancing risks of participating certain social activities can then be evaluated using different criteria for different social groups. Theoretical and simulation results for the distancing game in such heterogeneous populations are discussed in the paper, showing that the game can be played in almost the same form as in the general case: just use different risk-assessment functions for different social groups; and if the interactions among the individuals are restricted within their own groups, each social group would eventually find its own equilibrium strategy while the whole population approaches to the average one.

Third, a small number of individuals can be selected to act as leaders and the rest of the population as followers. The leaders make decisions on their own strategies while the followers simply copy the strategies of the leaders. The distancing game can then be carried out with such a leader-follower scheme. The simulation results show that the game can indeed proceed without requiring every individual to make rational decisions, and reach its equilibrium successfully even when only a small number of individuals, say 20% of the population, are designated as leaders. Indeed, in practice, it is most likely that only a small number of individuals such as public health experts or community leaders make decisions or recommendations while all others just follow.

As such, the work in this paper confirms the game model as a plausible approach to the study of the collective behavior of social distancing. It shows how the distancing game is played out in possible realistic as well as idealistic scenarios, reveals the conflicting yet cooperative nature of social distancing, and sheds lights on a self-organizing, bottom-up perspective of distancing practices.

Results

As a population game

Consider a population of $m$ individuals with $n$ social activities. Assume that every individual needs to decide a frequency to participate in each of the activities, say $x_i$ for activity $i$. Then the collection of these frequencies $x = \{x_i : i = 1, \ldots, n\}$ can be considered as a distancing strategy of the individual. Let $y_i$ be the average frequency of the population to participate in activity $i$. Then, the collection of these average frequencies $y = \{y_i : i = 1, \ldots, n\}$ can be considered as a distancing strategy of the population. Here a frequency $x_i$ or $y_i$ can be represented by the active hours in activity $i$ in a week (total 112 hours per week if 16 hours are counted as active hours per day excluding 8 hours sleeping time).

Given a distancing strategy $y$ of the population, assume that the potential distancing risk of having close social contacts and negative social/economic impacts in activity $i$, when fully participated, can be represented by a function $p_i(y)$. Then, the distancing risk of an individual of strategy $x$ at activity $i$ must be $x_i p_i(y)$, and at all the activities together be $\Sigma_i x_i p_i(y)$, where $\Sigma$ means the sum over all $i$'s. Let this summation be denoted as a function $\pi(x, y)$. A distancing game can then be defined for an individual against the population with $\pi(x, y)$ as the cost function; and a
strategy $x^*$ is an equilibrium strategy for the game if and only if every individual in the population takes this strategy (and hence $y^* = x^*$), and his/her distancing risk $\pi(x^*, y^*)$ using strategy $x^*$ is no greater than the distancing risk $\pi(x, y^*)$ using any other strategy $x$.

To make the matter simpler, assume that the social activities are independent of each other, i.e., the individuals participating in one of the activities do not have contacts with those in other activities. When the activities are independent, the function $p_i(y)$ can be defined to have two parts, one for close social contacts and the other for negative social/economic impacts, both proportional only to the participating frequency $y_i$ of the population in activity $i$. The amount of close social contacts can be estimated by a term $\alpha_i y_i$ with $\alpha_i$ named as the contact factor of activity $i$. The amount of negative social/economic impacts can be estimated by a term $\beta_i y_i$ with $\beta_i$ named as the impact factor of activity $i$. The unit for $\beta_i$ is the same as $\alpha_i$, i.e., the number of contacts per hour, but for $\alpha_i$ it is the number of close contacts one may have if fully participating in activity $i$, while for $\beta_i$ it is the number of necessary contacts one may lose when staying only in this activity. Table 1 gives a list of $\alpha_i$ and $\beta_i$ values assigned to 20 commonly attended social activities or CASA activities for short. They are estimated based on common practices or public guidelines such as the CDC social distancing recommendations [26] (details in Deriving contact and impact factors in Methods). They will be used throughout the paper for the purpose of concept proofing. In real applications, they certainly need to be further refined and justified with more experimental data.

### Table 1. Estimated contact factors and impact factors for 20 CASA activities

<table>
<thead>
<tr>
<th>Act:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_i$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>2.8000</td>
<td>2.8000</td>
<td>4.6667</td>
<td>4.6667</td>
<td>2.8000</td>
<td>4.6667</td>
<td>4.6667</td>
<td>2.8000</td>
<td>2.8000</td>
<td>4.6667</td>
</tr>
<tr>
<td>Act:</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>$\alpha_i$</td>
<td>3.5</td>
<td>3.5</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>4.6667</td>
<td>2.8000</td>
<td>3.5000</td>
<td>3.5000</td>
<td>0.8750</td>
<td>0.7000</td>
<td>2.3333</td>
<td>2.3333</td>
<td>3.5000</td>
<td>3.5000</td>
</tr>
</tbody>
</table>

**Table Legends:** Act – activity: 1 - reading or watching TV, 2 - working at home, 3 - hiking, 4 - gardening, 5 - stay with family, 6 - grocery shopping, 7 - go to hospitals, 8 - visit friends, 9 - cafeteria dinning, 10 - go to shopping malls, 11 - taking buses, 12 - going to church, 13 - watch sports, 14 - attend concerts, 15 - go to schools, 16 - go to workplaces, 17 - large gathering, 18 - go to bars or night clubs, 19 - air traveling, 20 - go to movie theaters; $\alpha_i$ – contact factors; $\beta_i$ – impact factors: #contacts lost per hour

By introducing another parameter $\delta_i$, the function $p_i(y)$ can be defined as $p_i(y) = \delta_i \alpha_i y_i + (1 - \delta_i) \beta_i y_i = w_i y_i$ with $w_i = \delta_i \alpha_i + (1 - \delta_i) \beta_i$ named as the risk factor of activity $i$. Here, $0 \leq \delta_i \leq 1$ can be used to balance between the risks of having close social contacts and negative social/economic impacts at activity $i$. Based on standard game theory, given $p_i(y) = w_i y_i$, if $w_i > 0$ for all $i$, an equilibrium strategy $x^*$ can be obtained with $x_i^* = \lambda / w_i$, $i = 1, \ldots, n$, where $\lambda$ is a constant such that $1/\lambda = \Sigma_i 1/w_i$ (more details in Equilibrium strategies and stabilities in Methods).

Table 2 gives the equilibrium strategies of the distancing game for the 20 CASA activities in Table 1 with $\delta_i = 1$, 0, or 0.5 for all $i$ corresponding to the cases for complete social distancing, no social distancing, or something in between, respectively. Fig 1 further illustrates the contrasts among these strategies, where for complete social distancing (blue line, $\delta_i = 1$), for the first 8 activities with small contact factors, the participating frequencies are high, while for the last 8 activities with large contact
factors, the participating frequencies are low. On the other hand, if no social distancing (red line, $\delta_i = 0$), for the first 8 activities, the participating frequencies are much lower than those for complete social distancing, while for the last 8 activities, the participating frequencies are much higher. When both social contacts and negative impacts are concerned (brown line, $\delta_i = 0.5$), for the first 8 activities, the participating frequencies are still higher than those with no social distancing, but not as high as those for complete social distancing, and for the last 8 activities, the participating frequencies are certainly lower than those with no social distancing, but not as low as those for complete social distancing.

Table 2. Equilibrium strategies for 20 CASA activities w/o social distancing

<table>
<thead>
<tr>
<th>Act:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_i = 1$:</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$\delta_i = 0.5$:</td>
<td>10.370</td>
<td>10.370</td>
<td>6.9537</td>
<td>6.9537</td>
<td>8.2092</td>
<td>5.9106</td>
<td>5.9106</td>
<td>8.2092</td>
<td>6.2546</td>
<td>4.8250</td>
</tr>
<tr>
<td>$\delta_i = 0$:</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Act:</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>$\delta_i = 1$:</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\delta_i = 0.5$:</td>
<td>4.8250</td>
<td>6.2546</td>
<td>3.7528</td>
<td>3.7528</td>
<td>5.0037</td>
<td>5.1174</td>
<td>2.4125</td>
<td>2.4125</td>
<td>2.2517</td>
<td>2.2517</td>
</tr>
<tr>
<td>$\delta_i = 0$:</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>20</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Table Legends: Act – activity: same as Table 1; $\delta_i = 1$ – complete social distancing; $\delta_i = 0$ – no social distancing; strategies represented by active times in hours per week (total 112 active hours)

Fig 1. The equilibrium strategies of the distancing game for complete social distancing (blue line), no social distancing (red line), and partially social distancing (brown line), displayed as active times in hours per week (total 112 active hours)

In small-world social networks

To be more realistic, assume that the population is distributed over a social network, and each individual only interacts with his/her neighbors in the social network. The distancing game can then be viewed as one played by every individual against the population in his/her neighborhood in the social network. Depending on the neighborhood size, the game becomes against a fraction of population ranging from the immediate neighbors of the individual to the whole population. For convenience, define the neighborhood of size $k$ of an individual to be one that includes all the neighbors connected to the individual with up to $k$ consecutive connections. Then, if $k$ is large enough, the neighborhood would include the whole population.

To mimic a real social network, the Watts-Strogatz algorithm \[89\] is used to generate a small-world social network. Assume that there are 200 individuals in the population ($m = 200$), the average degree of the nodes is 6 ($K = 6$), and 30 percent of all the links for each node come from random connections ($b = 0.3$). Fig 2 shows the generated network and the distribution of the degrees of the nodes. The nodes are displayed around a circle. About 70 percent of the links are along the edges which are not clearly visible in the graph. The rest of the links are more visible in the interior of the circle. In general, as the degree parameter $K$ or the randomness parameter $b$ increases, the graph becomes denser with more interior links (details in Generating small-world social networks in Methods).

Assume that every individual plays the distancing game with the population in a neighborhood of the same size. Let $x$ be the strategy of any individual and $y$ the
average strategy in the corresponding neighborhood. Then, the potential distancing risk at activity $i$ can still be estimated using the function $p_i(y)$, with which the individual can update his/her own strategy $x$: If $p_i(y)$ is higher than the average over all other activities, activity $i$ must be riskier, and $x_i$ in activity $i$ should be decreased; otherwise be increased. The update can be repeated for every individual in the population until no one can further improve his/her strategy, and the game hopefully reaches an equilibrium.

The distancing game is simulated on networks similar to the one in Fig 2. The simulation algorithm is outlined in Simulation of distancing games in social networks in Methods. The parameters for the networks are fixed to $m = 200$ and $K = 6$. The activities in Table 1 are used for the game and therefore, $n = 20$. Other parameters for the game are varied with the balancing parameter $b = 0.10, 0.20, 0.30, 0.40, 0.50$, and the neighborhood size $k = 1, 2, 3, 4, 5, 6$. For each set of parameters, an equilibrium strategy $x^*$ for every individual and $y^*$ for the population for the general distancing game (without network) can be computed directly as described in the previous section. Let $x$ be the strategy for an individual obtained by the simulation and $y$ the corresponding neighborhood strategy. These quantities are recorded in the simulation and compared with the equilibrium strategies $x^*$ and $y^*$. The results for all the parameter settings are documented (in Simulation results 1 in Supplementary Information).

Table 3 shows the simulation results for the game on the network in Fig 2 with $\delta_i = 0, 0.5, 1$ respectively for all $i$. The neighborhood size $k$ is changed from $1$ to $2, 3, 4, 5, 6$. Correspondingly, the average number of neighbors for each individual changes from $7$ to $28, 92, 179, 199, 200$. The last number agrees with the general consensus on small-world social networks where every pairs of individuals can be found connected with up to more or less 6 consecutive connections [90][92]. It is therefore not surprising that when $k = 6$, the game on this network converges to the same equilibrium strategy as the general distancing game (without network), for it is almost the same as the general distancing game played by every individual against the whole population. It is surprising, however, that when the neighborhood size is reduced, even when $k = 2$ with only 28 neighbors in average for each individual, the game still converges to the equilibrium strategy of the general distancing game, with $x$ converging to $x^*$ and $y$ to $y^*$. Only for $k = 1$, the game fails to converge, and the simulation terminates when not making any progress. These results are observed to be consistent for all other cases with varying $b$ and $\delta_i$ values (in Simulation results 1 in Supplementary Information).

In heterogeneous populations

Not every individual is equally vulnerable for epidemic infection. Nor is every individual equally likely to spread the disease. For example, for COVID-19, children seem not as susceptible to infection as adults, and they are less likely to carry and spread the virus. Similarly, the vaccinated people are more or less immune to infection and are probably more free to participate in social activities than those unvaccinated.

Further, the essential workers or economically vulnerable individuals are more concerned with the social/economic impacts than the health benefits of social distancing. A population should therefore be divided into different groups who rate the distancing risks of social activities differently.

As an example, consider a population evenly divided into 4 population groups:
Table 3. Convergence of distancing strategies in small-world social networks

<table>
<thead>
<tr>
<th>$\delta_i$</th>
<th>$k =$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>$</td>
<td></td>
<td>x - x^*</td>
<td></td>
<td>$</td>
<td>5.13e-03</td>
<td>5.76e-05</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td></td>
<td>y - y^*</td>
<td></td>
<td>$</td>
<td>2.71e-03</td>
<td>1.16e-05</td>
</tr>
<tr>
<td>0.50</td>
<td>$</td>
<td></td>
<td>x - x^*</td>
<td></td>
<td>$</td>
<td>2.74e-02</td>
<td>7.61e-05</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td></td>
<td>y - y^*</td>
<td></td>
<td>$</td>
<td>1.04e-03</td>
<td>1.49e-05</td>
</tr>
<tr>
<td>1.00</td>
<td>$</td>
<td></td>
<td>x - x^*</td>
<td></td>
<td>$</td>
<td>3.59e-03</td>
<td>7.99e-05</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td></td>
<td>y - y^*</td>
<td></td>
<td>$</td>
<td>1.97e-03</td>
<td>1.64e-05</td>
</tr>
</tbody>
</table>

Table Legends: $x$ – individual strategy from simulation; $y$ – neighborhood strategy from simulation; $x^*$ – individual strategy at equilibrium (without network); $y^*$ – population strategy at equilibrium (without network); $\parallel \cdot \parallel$ – the Euclidean norm: $\parallel x \parallel^2 = x_1^2 + \cdots + x_n^2$; $< >$ – average over all individuals.

either according to the ages of the individuals into $g_1$: 1-20 years old; $g_2$: 21-40; $g_3$: 41-60; and $g_4$: 61-80, or according to the level of protection of the individuals into $g_1$: the vaccinated; $g_2$: the recovered; $g_3$: the unprotected; and $g_4$: the most vulnerable.

Set $\delta_i$ for each of the groups to 0.00 for $g_1$, 0.25 for $g_2$, 0.75 for $g_3$, and 1.00 for $g_4$, thereby making the first two groups more open to expand their social activities but the last two to prefer more social distancing. Fig 3 demonstrates the results from computer simulation for the distancing game played among these population groups.

The simulation is conducted in a similar setting as for the game on the network shown in Fig 2 with the neighborhood size fixed to $k = 3$. In each generation of the simulation, every individual plays the game once, i.e., has a chance to update his/her strategy. However, different from the simulation described in the previous section, when evaluating the potential distancing risks, the contributions from different population groups are different due to their different $\delta_i$ values. When they are counted, more weight is also given to the contribution from the individual’s own group than from other groups. In addition, when comparing with the population average on the distancing risks, only the average over the individual’s own group is considered (details in Distancing in heterogeneous populations in Methods).

Fig 3 displays four snapshots from the simulation, showing the changes of the individual strategies of the game in four different generations. The circles represent the individual strategies and the stars the average population strategies. The circles are color coded for different population groups, with blue for $g_1$, cyan for $g_2$, magenta for $g_3$, and red for $g_4$. For each activity, there are 200 circles corresponding to the participating frequencies of 200 individuals for the activity. The labels on the $x$-axis are 20 CASA activities as defined in Table 1. The first graph in the figure shows the strategies of the individuals at the beginning of the 1st generation of the simulation, which are randomly generated around reasonably guessed starting strategies for each of the population groups. The second graph shows the strategies of the individuals after the 3rd generation, when they start separating into different groups. The third graph shows the strategies after the 6th generation, when they almost converge to their equilibrium positions. The last graph shows the strategies after the 9th generation, when they are close enough to their equilibrium values, and the simulation is terminated.

In the end of the simulation, each population group reaches an equilibrium strategy or more rigorously, its approximation. Since $\delta_i = 0.00$ and 0.25 for all $i$ for $g_1$ and $g_2$, the individuals in these two groups are considered to be more risk-taking, and their
Fig 3. The simulation results for an example distancing game with a heterogeneous population. Shown in the figure are the individual and population strategies of the game in four different generations of the simulation. The stars represent the population strategies and the circles the individual strategies. The circles are color coded for different population groups, with blue for $g_1$, cyan for $g_2$, magenta for $g_3$, and red for $g_4$. Over each activity, there are 200 circles corresponding to the strategies for 200 individuals. Labeled in the x-axis are 20 CASA activities. Along the y-axis are the active times in hours per week.

frequencies to participate in the socially active though high-contact activities (13-20) appear to be higher than the population average, while their frequencies to stay with the low-contact but socially isolating activities (1-8) are lower than the population average. On the other hand, since $\delta_i = 0.75$ and 1.00 for all $i$ for $g_3$ and $g_4$, the individuals in these two groups are considered to be more conservative, and their frequencies to stay with the low-contact though socially isolating activities (1-8) appear to be higher than the population average, while their frequencies to join the socially active but high-contact activities (13-20) are lower than the population average.

Simulations for distancing games with different population groups are conducted with varying balancing parameter $\delta_i$, randomness parameter $b$, and neighborhood size $k$. The results from these simulations (documented in Simulation results 2 in Supplementary Information) are all consistent with what are observed in the above example, showing that the game model can be extended to heterogeneous populations to predict different distancing behaviors among different population groups. What unexpected in these results is that the game for each population group converges to its own equilibrium strategy as if it is played alone, while the population strategy is simply a collective result of all these individual group strategies, which can also be justified in theory (details in Distancing in heterogeneous populations in Methods). This may not be surprising, on the other hand, as observed during the COVID-19 pandemic that children and adults do have their own distancing strategies as their shares in the average distancing strategy of the population.

By following the leaders

Not every individual actively participates in social distancing. Even if he/she does, he/she may not necessarily make the decisions as accurately as assumed such as evaluating the distancing risks of the activities and responding with appropriate actions, etc. In practice, it is most likely that only a small number of individuals such as public health experts or community leaders make decisions or recommendations while all others just follow. Indeed, leadership plays an important role in collective actions in both nature and human societies [93–97].

In order to incorporate the leadership factor into the distancing activities, a small number of individuals are designated as leaders and the rest of the population as followers. A leader makes a distancing decision as a regular player in the distancing game, while a follower just copies the strategies of some leaders unless he/she cannot find a leader in his/her neighborhood when he/she either makes his/her own decision or simply follows the crowd (details in Following the leaders vs. following the crowd in Methods).

The game with mixed leaders and followers is simulated in a small-world social network as given in Fig 2 with varying neighborhood sizes and percentages of leaders in the population. It is also assumed to be against a heterogeneous population as given in the example game in the previous section, where there are four population
groups, and the first two groups contribute to the distancing risks differently from the last two. Table 4 contains some of the simulation results with the neighborhood size $k = 3$ but varying percentages of leaders in the population.

### Table 4. Convergence of distancing strategies in mixed leader-follower populations

<table>
<thead>
<tr>
<th>% leaders</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>60%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_1$: $|x - x^*|$</td>
<td>1.97e-02</td>
<td>1.70e-02</td>
<td>1.58e-02</td>
<td>1.09e-02</td>
<td>1.15e-02</td>
<td>8.36e-03</td>
</tr>
<tr>
<td>$g_2$: $|x - x^*|$</td>
<td>1.03e-02</td>
<td>1.17e-02</td>
<td>1.04e-02</td>
<td>7.48e-03</td>
<td>3.84e-03</td>
<td>2.42e-03</td>
</tr>
<tr>
<td>$g_3$: $|x - x^*|$</td>
<td>1.41e-02</td>
<td>1.26e-02</td>
<td>9.07e-03</td>
<td>8.10e-03</td>
<td>6.47e-03</td>
<td>2.62e-03</td>
</tr>
<tr>
<td>$g_4$: $|x - x^*|$</td>
<td>1.38e-02</td>
<td>1.29e-02</td>
<td>1.06e-02</td>
<td>7.98e-03</td>
<td>1.01e-03</td>
<td>5.87e-03</td>
</tr>
</tbody>
</table>

### Table Legends:
- $x$ – individual strategy from simulation;
- $y$ – group strategy in neighborhood from simulation;
- $x^*$ – individual strategy at equilibrium;
- $y^*$ – group strategy at equilibrium;
- $<1>$ – results by following the crowd if leaders not found;
- $<2>$ – results by self-determination if leaders not found.

The first set of results $<1>$ in the table is obtained with a follow-the-crowd strategy if a follower cannot find a group leader among his/her closest neighbors. When there is a high percentage of leaders in the population ($\geq 40\%$), the game converges well to the equilibrium strategies for all the individual groups. If there are lower than 20\% leaders in the population, the convergence becomes less accurate (see top two plots in Fig 4), when most of the followers are not able to find a group leader in their neighborhood, and there is a disadvantage by simply following the crowd.

The second set of results $<2>$ is obtained with a make-own-decision strategy if a follower cannot find a group leader among his/her closest neighbors. Similar to the previous case, when there is a high percentage of leaders in the population ($\geq 40\%$), the game converges well to the equilibrium strategies for all the individual groups. In contrast to the previous case, when the percentage decreases to $\leq 20\%$, the convergence remains to be as accurate (see bottom two plots in Fig 4), because when there are fewer leaders, more followers start making their own decisions, which can be even better than following the leaders.

### Fig 4. The convergence results for the distancing games with leaders. The difference between individual strategy $x$ and equilibrium strategy $x^*$ is computed and averaged over all the individuals, and plotted at each generation. The top two graphs from left to right are the plots for the game with 40\% and 20\% of the individuals being assigned as leaders, and a follow-the-crowd strategy is used if a follower cannot find a group leader among his/her closest neighbors. The bottom two graphs from left to right are the plots for the game with 40\% and 20\% of individuals being assigned as leaders, and a make-own-decision strategy is used if a follower cannot find a group leader among his/her closest neighbors.

The simulation results with other neighborhood sizes and percentages of leaders (documented in Simulation results 3 in Supplementary Information) are all consistent with those in the above example. In general, when the percentage of leaders in the
population is high, the game is expected to perform well and reach its equilibrium strategy. When the percentage is small, the game still runs reasonably well, showing that a small group of leaders are able to guide, and not every individual is required to make a decision. However, as shown in contrast between the set <1> and set <2> results, when there are fewer leaders, making own decisions actively is certainly more reliable than simply following the crowd.

Discussion

The social distancing activities are not easy to track and hence difficult to study experimentally. The model proposed in this work presents a theoretical framework with which the distancing behaviors can be simulated, predicted, and analyzed. In this model, every individual in a given population is assumed to engage in social distancing by playing a distancing game, where based on what everybody else does, every individual makes decision on how to participate in a given set of social activities so that he/she can minimize his/her close social contacts with least possible negative social/economic impacts.

A distancing decision can be made to focus more on reducing close social contacts ($\delta_i > 0.5$) or minimizing negative social/economic impacts ($\delta_i < 0.5$). It depends on the severity of the epidemic and how the population responds to it. In general, if the epidemic is in a severe stage, a relatively large $\delta_i$ value would be good to reduce close social contacts, although the negative social/economic impacts may be high. If instead, a relatively small $\delta_i$ value is adopted, the amount of negative social/economic impacts is controlled, but the population would be exposed to more close social contacts. For example, for the distancing game demonstrated in Fig 1 in Results – As a population game when in a severe epidemic, if $\delta_i$ is set to 1 for all $i$, the amount of close contacts for every individual is reduced to about 14 per week at equilibrium with the lost amount of necessary contacts around 35.21. However, if $\delta_i$ is set to 0 for all $i$ instead, the amount of close contacts for every individual is increased to 59.95 per week, although the lost amount of necessary contacts is minimized to 14 per week at equilibrium.

The general distancing game assumes that every individual interacts with all others in the population. To be realistic, in this work, the population is assumed to be distributed over a social network, and each individual interacts only with his/her social neighbors. An algorithm is implemented to simulate the distancing game on such a network as discussed in Results – In small-world social networks. The algorithm is not equivalent to the general distancing game, and is not guaranteed to converge to an equilibrium strategy, either. In addition, the distancing game on a social network is played by every individual against his/her neighborhood, not the whole population. Nonetheless, the games simulated on social networks all converge to the equilibrium strategies of the general distancing games even when the neighborhood sizes are small and are restricted to contain only close neighbors, which is surprising, or not, as it may be how it is played out in the real world.

The extension of the distancing game to populations with multiple population groups in Results – In heterogeneous populations allows each population group to have its own assessment on distancing risks and hence its own distancing strategy. The equilibrium strategies of such games are justified in theory and also by simulation. In these games, the individuals take into account the contributions to their distancing risks from all population groups while giving more weights to those from their own groups, as they are supposed to interact more with the individuals in their own groups than those in other groups. However, the individuals may have more interactions with the individuals in some other groups as well as shown in several recent
publications [98,99], suggesting that in general, heavier weights may be assigned across more than one groups based on possible group-group interaction patterns.

The leadership role in social distancing is addressed lightly in this work as discussed in [Results]. By following the leaders. In fact, the leadership in social distancing is way beyond a matter of a few leaders making their own distancing decisions. For better or worse, leadership is often a determining factor in directing or changing the social distancing activities, as local or global organizations or governments make public health policies, provide social distancing guidelines, or give lockdown orders, etc., which are out of the scope of this study. However, the work in this study may help to understand the nature of social distancing as a collective behavior of human population, thereby providing a quantitative approach to assessing and improving the outcomes of public health policies concerning the control of social activities and its potential impacts.

The games discussed in this work are assumed to have only a small number of very general social activities, i.e., 20 CASA activities. In practice, there can be many more activities. They can also be of more specific types. For example, there can be different workplaces, different restaurants, and different shopping centers, and going to each of these places may be considered as a different social activity. The activities may also have some connections, i.e., not necessarily be independent of each other. An individual who participating in one of the activities may have contacts with individuals in other activities when the activities are carried out in close proximity in time or space. However, the proposed model can be extended to dependent activities by including the dependency among the activities in the estimation of the distancing risks [86].

The number of population groups in a heterogeneous population is not limited either, although only up to four population groups are considered in this work. The types of groups can also be combined. For example, groups can be formed according to the age as well as the vulnerability to the disease such as the vaccinated/unvaccinated children, recovered/unprotected adults, etc. In any case, the distancing activities in such populations can all be modeled as a multi-player game with each group corresponding to a single player and having its own risk assessment. In addition, in this work, the parameter $\delta_i$ is always set to a given value for all $i$, but in practice, it can be given a different value for a different $i$. For example, in certain situations, close contacts in schools or workplaces are less of a concern, and $\delta_i$ for these activities may therefore be given a smaller value than it is supposed to be. Further investigation into the use of this variation in practice can be pursued in future efforts.

Methods

Deriving contact and impact factors

The contact factor $\alpha_i$ in function $p_i$ is the number of possible close contacts per unit time one may have in activity $i$ when the activity is fully attended. It is not easy to estimate in general, for it may change from place to place or from time to time for a specific activity. In some sense, it is even unclear what a close contact means and how to measure it. In this work, only a general set of 20 CASA activities is considered. Their contact factors are determined by solving a so-called inverse game or inverse problem as demonstrated below.

First, the 20 CASA activities are grouped to 5 levels according their possible contact rates: The lowest level – activities 1-4; the second level – activities 5-8; the third level – activities 9-12; the fourth level – activities 13-16; and the fifth level – activities 17-20, according to common practices and CDC recommendations [26].
each level of activities, a contact factor is to be assigned. Then, note that when \( \delta_i = 1 \)
for all \( i \), the distancing game is equivalent to one for complete social distancing.
Assume that the equilibrium strategy \( x^* \) and hence \( y^* \) in such a situation can be
estimated. For example, with complete social distancing, an individual may stay home
almost the whole day every day; he/she may spend 2 hours per day (or equivalently,
14 hours per week) for each of the first level activities, 1 hour per day (or equivalently,
7 hours per week) for each of the second, and so on and so forth (as listed in Table 5).
Then, the contact factor \( \alpha_i \) can be determined based on the condition \( \alpha_i y^* = \lambda \)
for some constant \( \lambda \) for all \( i \). Set \( \alpha_i \) to 1 for the first level activities to obtain \( \lambda = 1/8 \)
contacts per hour (or equivalently, 14 contacts per week). The values for the rest of
the \( \alpha_i \) factors can then be extracted as given in Table 1.

Table 5. Estimated active times in 20 CASA activities for complete social distancing

<table>
<thead>
<tr>
<th>Act:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time:</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Act:</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time:</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table Legends: Act – activity: same as Table 1; Time – active time in hours per week (112 active hours)

Similarly, when \( \delta_i = 0 \) for all \( i \), the distancing game is equivalent to one where the
individuals are free to choose their social activities as if there is no epidemics and no
social distancing. Assume that the equilibrium strategy \( x^* \) and hence \( y^* \) in such a
situation can be estimated. For example, in a week, if without social distancing, in
average, an individual would work at home for 5 hours, stay with family for 5 hours,
visit friends for 4 hours, and go to workplaces for 20 hours, etc. (as listed in Table 6).
Then, the impact factor \( \beta_i \) can be determined based on the condition \( \beta_i y^*_i = \lambda \)
for some constant \( \lambda \) for all \( i \). Choose \( \lambda \) to be the same level as for complete social
distancing, i.e., \( \lambda = 1/8 \) contacts per hour (or equivalently, 14 contacts per week).
Then, The values for all \( \beta_i \) can be extracted as given in Table 1.

Table 6. Estimated active times in 20 CASA activities when free of epidemics

<table>
<thead>
<tr>
<th>Act:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time:</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Act:</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time:</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>20</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Table Legends: Act – activity: same as Table 1; Time – active time in hours per week (112 active hours)

Note that the active times assumed for the 20 CASA activities in Table 5 and
Table 6 are based on common distancing practices. They are used to estimate the
values for \( \alpha_i \) and \( \beta_i \), which then define a plausible test case required for the analysis
and simulation in this work. In real applications, they certainly need to be further
refined and justified with more experimental data. Note also that the unit for \( \beta_i \) is the
same as \( \alpha_i \), i.e., the number of contacts per hour, but for \( \alpha_i \) it is the number of close
contacts one may have when fully participating in activity \( i \), while for \( \beta_i \) it is the
number of necessary contacts one may lose when staying only in this activity.
Equilibrium strategies and stabilities

For each activity $i$, $p_i(y) = w_i y_i$ with $w_i = \delta_i \alpha_i + (1 - \delta_i) \beta_i$ and $0 \leq \delta_i \leq 1$. If $w_i > 0$ for all $i$, the game reaches equilibrium when the potential distancing risks at all the activities are the same, i.e., an optimal strategy $x^*$ and hence $y^*$ is found such that $p_i(y^*) = w_i y_i^* = \lambda$ for some constant $\lambda$ (proofs in Appendix A1: General distancing games). Note that $y^*_i > 0$ for all $i$, implying that the number of active activities is kept to the maximum ($= n$) at equilibrium.

Note also that the equilibrium strategy $x^*$ and hence $y^*$ is evolutionarily stable – a term used in evolutionary game theory [59,60]. It means that if there is a small change in the strategy, the equilibrium strategy still prevails. In other words, if the population strategy $y^*$ is perturbed (or invaded) slightly by a new strategy $y$, $y^*$ will remain to be a better choice than $y$, and not be taken over by $y$ (proofs in Appendix A2: Evolutionary stability).

Generating small-world social networks

The social network is generated with the well-known Watts-Strogatz algorithm [89]. The algorithm has three parameters to determine a social network, $m$ the number of the nodes, $K$ the average degree of the nodes, and $b$ the randomness of the connections. The algorithm generates a small-world social network of $m$ nodes for a population of $m$ individuals. In a cyclic order, the algorithm first connects each node with $K/2$ nodes next to the node on the right and then on the left. Then, for each node $i$ and node $j$ of $K/2$ nodes connected to node $i$ on the right, the algorithm selects a node $k$ not connected to $i$, and with a probability $b$, removes the link between $i$ and $j$ and connects $i$ and $k$. In this way, the average degree of the nodes in the network would be around $K$, and the randomness of the connection between the connected nodes can be specified by $b$. A detailed algorithmic description for generating a small-world social network is given in Algorithm 1. A Matlab code can be found in the provided Simulation code 1, 2, 3 in Supplementary Information.

Algorithm 1 Generate a small-world network: $(m, K, b)$

| Require: $m$ integer $\lor K$ even $\lor 0 \leq b \leq 1$ |
| Ensure: Set $m$ nodes in a cycling order |
| 1: For $i = 1 : m$ |
| 2: Connect $i$ with $K/2$ nodes on its left |
| 3: Connect $i$ with $K/2$ nodes on its right |
| 4: End |
| 5: For $i = 1 : m$ |
| 6: For each $j$ of $K/2$ nodes next to $i$ |
| 7: If $(i,j)$ connected |
| 8: Find a node $k$ not connected with $i$ |
| 9: Disconnect $(i,j)$ and connect $(i,k)$ with a probability $b$ |
| 10: End |
| 11: End |
| 12: End |

Simulation of distancing games in social networks

The simulation of distancing games is based on the general principle of replicator dynamics for population games [59,60]. If $y_i$ is the participating frequency of the
population in activity $i$ at a certain time $t$, the replicator dynamics states that the changing rate of $y_i$ is proportional to the difference between the potential distancing risk $p_i(y)$ at activity $i$ and the population average $\Sigma y_i p_i(y)$. If the potential distancing risk is higher than the average, activity $i$ is considered to be riskier, and $y_i$ should be increased; otherwise, $y_i$ should be decreased.

For the game on a social network, the simulation can be done for every individual in his/her neighborhood. The population strategy $y$ then becomes the neighborhood strategy, which is different for a different individual in general. The neighborhood of size $k$ of an individual includes all the neighbors connected to the individual with up to $k$ consecutive connections. Depending on the neighborhood size, the game becomes against a fraction of population ranging from the immediate neighbors of the individual to the whole population. Algorithm 2 gives more algorithmic description on how an individual updates his/her strategy in every round of the game. A Matlab code for the whole simulation is provided (in Simulation code 1 in Supplementary Information).

The simulation starts with a strategy for every individual randomly generated around its equilibrium one. More specifically, if $x_i^*$ is the equilibrium value of the participating frequency for activity $i$, then first perturb $x_i^*$ randomly by a certain percentage, say 20%, and then generate $x_i$ randomly within 80% of deviation from the perturbed value of $x_i^*$. The simulation proceeds in multiple generations. At each generation, every individual plays the game once, i.e., has a chance to update his/her strategy. The simulation ends when either every individual strategy converges to the equilibrium strategy in average or it stops making any further progress. Every simulation is repeated for 5 times and an average output is recorded and reported.

### Distancing in heterogeneous populations

Consider a simple case where the population is divided into two groups, groups $a$ and $b$. Let $x^a$ and $x^b$ be the distancing strategies for individuals in groups $a$ and $b$, respectively, with $x_i^a$ and $x_i^b$ being the participating frequencies of the individuals in activity $i$. Let $y^a$ and $y^b$ be the average strategies of the individuals in groups $a$ and $b$ in the population, with $y_i^a$ and $y_i^b$ being the corresponding average participating frequencies of these individuals in activity $i$. Given strategies $y^a$ and $y^b$ in the population, the potential distancing risk at activity $i$ can be estimated by a function $p_i^a(y^a, y^b) = (1 + s) w_i^a y_i^a + w_i^b y_i^b$ for a group $a$ individual or $p_i^b(y^a, y^b) = w_i^a y_i^a + (1 + s) w_i^b y_i^b$ for a group $b$ individual, where $w_i^a$ and $w_i^b$ are risk factors for groups $a$ and $b$, respectively, and the contribution to the distancing risk from the individual's own group is given more weight $(1 + s)$ for some $s > 0$, as the individual is supposed to interact more with the individuals in his/her own group than those in the other group.

Then, for an individual of strategy $x^a$ in group $a$, the distancing risk to participate in given $n$ activities can be evaluated by a function $\pi(a, y^a, y^b) = \sum_i x_i^a p_i^a(y^a, y^b)$. Similarly, for an individual of strategy $x^b$ in group $b$, the distancing risk to participate in given $n$ activities can be evaluated by a function $\pi_b(x^b, y^a, y^b) = \sum_i x_i^b p_i^b(y^a, y^b)$. Together, with these functions, a multi-player distancing game can be defined for the whole population with each population group corresponding to a single player; and a pair of strategies $x^{a*}$ and $x^{b*}$ form an equilibrium pair of strategies for the game if and only if $y^{a*} = x^{a*}$ and $y^{b*} = x^{b*}$, and for every individual in group $a$, the distancing risk $\pi(a, y^{a*}, y^{b*})$ using strategy $x^{a*}$ is no greater than the distancing risk $\pi(a, y^{a*}, y^{b*})$ using any other strategy $x^a$, and for every individual in group $b$, the distancing risk $\pi_b(x^{b*}, y^{a*}, y^{b*})$ using strategy $x^{b*}$ is no greater than the distancing risk $\pi_b(x^{b*}, y^{a*}, y^{b*})$ using any other strategy $x^b$.

It follows from a little bit analysis that the equilibrium strategy for each
Algorithm 2 Updating individual distancing strategies

Require: Individual strategy $x$, neighborhood strategy $y$

1: $\bar{p}_y = \sum_i y_i p_i(y)$
2: For $i = 1 : n$
3:   If $p_i(y) < \bar{p}_y$
4:     If $x_i < y_i$
5:       $x_i = x_i + 0.9 \times (y_i - x_i)$
6:     Else
7:       $x_i = x_i + 0.1 \times \min\{1 - x_i, x_i - y_i\}$
8:   End
9: End
10: If $p_i(y) > \bar{p}_y$
11:   If $x_i > y_i$
12:     $x_i = x_i - 0.9 \times (x_i - y_i)$
13:   Else
14:     $x_i = x_i - 0.1 \times \min\{x_i, y_i - x_i\}$
15: End
16: End
17: If $|p_i(y) - \bar{p}_y| < 0.01$
18:   If $x_i > y_i$
19:     $x_i = x_i - 0.5 \times (x_i - y_i)$
20: End
21:   If $x_i < y_i$
22:     $x_i = x_i + 0.5 \times (x_i - y_i)$
23: End
24: End
25: $x = x / \sum_i x_i$

population group can be obtained with $x_i^a = \lambda_a / w_i^a$ for all $i$ for some constant $\lambda_a$ such that $1 / \lambda_a = \sum_i 1 / w_i^a$, and $x_i^b = \lambda_b / w_i^b$ for all $i$ for some constant $\lambda_b$ such that $1 / \lambda_b = \sum_i 1 / w_i^b$. At equilibrium, $y_i^a = x_i^a$ and $y_i^b = x_i^b$, and the average population strategy should be $y^a = \rho_a y_i^a + \rho_b y_i^b$, where $\rho_a$ and $\rho_b$ are the percentages of group $a$ and $b$ individuals in the population, respectively. These results can be extended straightforwardly to populations with more than two population groups (general descriptions and proofs in Appendix A3: Games with multiple population groups).

The simulation of the distancing game with multiple population groups is done with the population also distributed over a small-world social network, where the game is played by every individual against his/her population group in his/her neighborhood. The key difference of this simulation from the one described in Algorithm 2 is that the potential distancing risk at each activity is estimated using a formula as described above, and the average risk over all activities is evaluated for each individual using his/her group strategy in his/her neighborhood. A Matlab code for simulating the distancing games with up to four population groups is provided (in Simulation code 2 in Supplementary Information).

Following the leaders vs. following the crowd

A certain percentage of individuals are randomly selected as leaders. A leader makes distancing decisions as a regular player for the distancing game, whether the game is
played in a small-world social network or with multiple population groups. A follower tries to find the leaders in his/her population group in his/her neighborhood, and copies the average strategy of the leaders. If he/she fails to find a leader among his/her closest neighbors, he/she either makes her own decision as a regular player or follows the crowd by copying the average strategy of his/her group members in his/her neighborhood. The simulation is done with the percentage of leaders in the population varying from 10% to 20%, 30%, 40%, 50%, and 60%, and the neighborhood size changing from 1 to 2, 3, 4, 5, and 6. The parameters for the network are fixed to $m = 200$, $K = 6$, and $b = 0.3$. A Matlab code for the simulation is provided (in Simulation code 3 in Supplementary Information).

**Supplementary information**

Results from computer simulation conducted in this work are documented in the following files. The Matlab codes producing the results are all provided.

- Simulation results 1: SI-simulation-results-1.pdf
- Simulation results 2: SI-simulation-results-2.pdf
- Simulation code 1: run_simulation_1.m; simulation_1.m
- Simulation code 2: run_simulation_2.m; simulation_2.m
- Simulation code 3: run_simulation_3.m; simulation_3.m

**Acknowledgments**

The author would like to acknowledge the support from the Simons Foundation on this work through the Mathematics and Physical Sciences Collaboration Grants for Mathematicians (Award Number: 586065).

**Appendix A**

**Appendix A1: General distancing games**

Assume that the population has $n$ activities. Let $x = \{x_i : i = 1, \ldots, n\}$ be a set of frequencies representing the distancing strategy of any individual, with $x_i$ being the frequency of the individual to participate in activity $i$, and $\Sigma_i x_i = 1$. Let $y = \{y_i : i = 1, \ldots, n\}$ be a set of frequencies representing the strategy of the population, with $y_i$ being the average frequency of all the individuals in the population to participate in activity $i$, and $\Sigma_i y_i = 1$.

Given a distancing strategy $y$ from the population, assume that each individual can estimate the potential distancing risk at each activity $i$ using a function $p_i(y)$. Then, the distancing risk of the individual of strategy $x$ at activity $i$ must be $x_i p_i(y)$, and at all the activities together be $\Sigma_i x_i p_i(y) = \pi(x, y)$.

**Definition 1** (Distancing game). A distancing game is a population game where every individual chooses a strategy $x$ against a strategy $y$ of the population so that his/her distancing risk $\pi(x, y)$ can be minimized.

**Definition 2** (Equilibrium strategy). A strategy $x^*$ is an equilibrium strategy of the distancing game if and only if every individual in the population takes this strategy $x^*$ (and hence $y^* = x^*$) such that his/her distancing risk $\pi(x^*, y^*) \leq \pi(x, y^*)$ for any strategy $x$. 

October 11, 2023
Theorem 1. A strategy \( x^* \) is an equilibrium strategy for the distancing game if and only if there is a constant \( \lambda \) such that
\[
    x_i^* (p_i(y^*) - \lambda) = 0, \quad x_i^* \geq 0, \quad p_i(y^*) - \lambda \geq 0, \quad i = 1, \ldots, n.
\]

Proof. \((\Rightarrow)\) Suppose that \( x^* \) is an equilibrium strategy. Then, \( \pi(x^*, y^*) \leq \pi(x, y^*) \) for any strategy \( x \), and therefore, \( \pi(x^*, y^*) \leq \pi(e_i, y^*) = p_i(y^*) \), \( i = 1, \ldots, n \), where \( e_i \) is the \( i \)th unit vector. Let \( \pi(x^*, y^*) = \lambda \). Then \( p_i(y^*) - \lambda \geq 0 \) for all \( i = 1, \ldots, n \). For any \( i \), if \( x_i^* = 0 \), then \( p_i(y^*) - \lambda = 0 \); if \( x_i^* > 0 \), \( p_i(y^*) - \lambda \) must be zero, for otherwise, \( x_i^*(p_i(y^*) - \lambda) > 0 \). Collect the latter inequality for all \( i \) to obtain \( \Sigma_i x_i^*(p_i(y^*) - \lambda) > 0 \). Then \( \pi(x^*, y^*) - \lambda > 0 \), which is contradictory to the fact that \( \pi(x^*, y^*) = \lambda \). Thus the conditions in (1) are all satisfied.

\((\Leftarrow)\) Suppose there is a parameter \( \lambda \) such that \( x^* \) satisfies all the conditions in (1).
Collect the first equation in (1) for all \( i \) to obtain \( \Sigma_i x_i^*(p_i(y^*) - \lambda) = 0 \), which is equivalent to \( \pi(x^*, y^*) - \lambda = 0 \), and therefore, \( \lambda = \pi(x^*, y^*) \). Let \( x \) be an arbitrary strategy. Multiply the last equation in (1) by \( x_i \) to obtain \( x_i(p_i(y^*) - \lambda) \geq 0 \). Collect the latter inequality to obtain \( \Sigma_i x_i(p_i(y^*) - \lambda) \geq 0 \), which is equivalent to \( \pi(x, y^*) - \lambda \geq 0 \). Then, \( \pi(x^*, y^*) \leq \pi(x, y^*) \) for any strategy \( x \), and \( x^* \) is an equilibrium strategy.

Theorem 2. Assume that the activities are independent and function \( p_i(y) = w_i y_i \) with \( w_i > 0 \) for all \( i \). Then, \( x_i^* = \lambda/w_i, \ i = 1, \ldots, n \), form an equilibrium strategy for the distancing game, where \( \lambda \) is a constant such that \( 1/\lambda = \Sigma_i 1/w_i \).

Proof. Let \( S^* = \{ i : x_i^* > 0 \} \). Then, by Theorem 1, for \( i \in S^* \), \( p_i(y^*) - \lambda = 0 \) for some constant \( \lambda \). It follows that \( w_i y_i^* = w_i x_i^* = \lambda \), and \( x_i^* = \lambda/w_i \) for all \( i \in S^* \). Since the sum of all \( x_i^* \) equals 1, the sum of the latter equations gives \( 1/\lambda = \Sigma_{i \in S^*} 1/w_i \). Note that \( S^* \) must contain all \( i = 1, \ldots, n \), for otherwise, if there is \( i \) such that \( x_i^* = 0 \), then \( p_i(y^*) = w_i y_i^* = w_i x_i^* = 0 \); it follows that \( p_i(y^*) < \lambda \), which contradicts to the fact that \( p_i(y^*) \geq \lambda \) for all \( i \) by Theorem 1.

Appendix A2: Evolutionary stability

Definition 3. Evolutionary stability. An equilibrium strategy \( x^* \) for the distancing game is evolutionarily stable if for any strategy \( x \neq x^* \), there is \( \epsilon \in (0, 1) \) such that \( \pi(x^*, \epsilon x + (1 - \epsilon)x^*) < \pi(x, \epsilon x + (1 - \epsilon)x^*) \) for all \( \epsilon \in (0, \epsilon) \).

Definition 4. Potential minimization. Let \( f(y) \) be a function such that \( f_i'(y) = p_i(y), \ i = 1, \ldots, n \). Then, the problem
\[
\min \{ f(y) : \Sigma_i y_i = 1, \ y_i \geq 0, \ i = 1, \ldots, n \}
\]

is called a potential minimization problem for the distancing game defined by \( p_i(y), \ i = 1, \ldots, n \).

Theorem 3. A strategy \( x^* \) is an equilibrium strategy for the distancing game if and only if \( x^* \) is a KKT point of the corresponding potential minimization problem.

Theorem 4. An equilibrium strategy \( x^* \) for the distancing game is evolutionarily stable if and only if \( x^* \) is a strict local minimizer of the corresponding potential minimization problem.

Theorem 5. Assume that the activities are independent and function \( p_i(y) = w_i y_i \) with \( w_i > 0 \) for all \( i \). Then, the equilibrium strategy \( x_i^* = \lambda/w_i, \ i = 1, \ldots, n \), for the distancing game is evolutionarily stable with \( 1/\lambda = \Sigma_i 1/w_i \).

Proof. The Hessian of the objective function \( f(x) \) of the potential minimization problem corresponding to the distancing game is a diagonal matrix with \( w_i \), \( i = 1, \ldots, n \) as the diagonal elements. Since \( w_i > 0 \) for all \( i \), the Hessian is positive.
definite, which guarantees the solution to the potential minimization problem $x^*$ to be a strict local minimizer. It follows that $x^*$ must be evolutionarily stable by Theorem [4].

**Appendix A3: Games with multiple population groups**

Assume that the population is divided into $M$ groups. Let $x^{(j)}$ be the strategy of an individual in group $j$, and $y^{(j)}$ the average strategy of all group $j$ individuals in the population. Let $w^{(j)}_i$ be the risk factor for activity $i$ for the individuals in group $j$. Then, the potential distancing risk for a group $j$ individual at activity $i$ can be defined as

$$p^{(j)}_i(y^{(j)}, y^{(-j)}) = \sum_k s_{jk} w^{(k)}_i \eta^{(k)}_i$$

where $y^{(-j)}$ represents all group strategies $y^{(1)}, \ldots, y^{(M)}$ excluding $y^{(j)}$, $\Sigma_k$ means the sum over all $k = 1, \ldots, M$, and $s_{jk}$ is a scaling factor, $s_{jk} = (1 + s)$ for some $s > 0$ if $k = j$ and $s_{jk} = 1$ if $k \neq j$, thus the contribution of group $j$ to the distancing risk is given more weight as an individual in group $j$ is supposed to interact more with the individuals in his/her own group than those in other groups.

**Definition 5** (Distancing game with multiple population groups). Assume that the population is divided into $M$ groups. Let $\pi_j(x^{(j)}, y^{(j)}, y^{(-j)}) = \sum_i x^{(j)}_i p^{(j)}_i(y^{(j)}, y^{(-j)})$, $j = 1, \ldots, M$. Then together, with all these functions, a multi-player distancing game can be formed with each population group corresponding to a single player; and a set of strategies $x^{(j)*}$, $j = 1, \ldots, M$, is an equilibrium set of strategies for the game if and only if for all $j = 1, \ldots, M$, $y^{(j)*} = x^{(j)*}$, and

$$\pi_j(x^{(j)*}, y^{(j)*}, y^{(-j)*}) \leq \pi_j(x^{(j)}, y^{(j)*}, y^{(-j)*})$$

for any strategy $x^{(j)}$.

**Theorem 6.** Assume that the population is divided into $M$ groups and the activities are independent. Assume that the function for an individual in group $j$ to evaluate the potential distancing risk at activity $i$ is given by (4) with $w^{(j)}_i > 0$ for all $i$ and $j$. Then, there is a unique set of equilibrium strategies $x^{(j)*}$, $j = 1, \ldots, M$, for the multi-player distancing game of the population, $x^{(j)*}_i = \lambda^{(j)} w^{(j)}_i$ for all $i$, where $\lambda^{(j)}$ is a constant such that $1/\lambda^{(j)} = \sum_i 1/w^{(j)}_i$.

**Proof.** Let $x^{(j)*}_i$ be the strategy of an individual in group $j$ at equilibrium and $y^{(j)*}_i$ the average strategy of group $j$ individuals in the population. Then, it is necessary and sufficient that for each group $j$, $p^{(j)}_i(y^{(j)*}_i, y^{(-j)*}_i) = t_j$ for all $i$ for some constant $t_j$, i.e.,

$$s_{11} w^{(1)}_i y^{(1)*}_i + \cdots + s_{1j} w^{(j)}_i y^{(j)*}_i + \cdots + s_{Mj} w^{(M)}_i y^{(M)*}_i = t_j$$

$$i = 1, \ldots, n; \quad j = 1, \ldots, M.$$  

The above equations can be written in a more compact form as:

$$Sz_i = t, \quad i = 1, \ldots, n$$

where $S = (s_{jk})$ is an $M \times M$ matrix, $s_{jk} = (1 + s)$ for some $s > 0$ if $j = k$ and $s_{jk} = 1$ if $j \neq k$, $z_i$ and $t$ are $M$-dimensional vectors, $z_i^T = (w^{(1)}_i y^{(1)*}_i, \ldots, w^{(M)}_i y^{(M)*}_i)$, and $t^T = (t_1, \ldots, t_M)$. It is not difficult to verify that $S$ is nonsingular. Therefore, $z_i = S^{-1}t$. Let $(S^{-1}t)_j = \lambda^{(j)}$, $j = 1, \ldots, M$. Then, $w^{(j)}_i y^{(j)*}_i = \lambda^{(j)}$ for all $i$. It follows that $y^{(j)*}_i = \lambda^{(j)}/w^{(j)}_i$ with $1/\lambda^{(j)} = \sum_i 1/w^{(j)}_i$.  

\[ \square \]
References


24. K. Pearce, What is social distancing and how can it slow the spread of COVID-19?, HUB, Johns Hopkins University, 2020, https://hub.jhu.edu/2020/03/13/what-is-social-distancing


55. G. Miller, Social distancing prevents infections, but it can have unintended consequences, Science News, March 16, 2020, https://doi.org/10.1126/science.abb7506


58. C. E. Kontokosta, B. Hong, and B. J. Bonczak, Measuring sensitivity to social distancing behavior during the COVID-19 pandemic, Nature Scientific Reports, 2022, https://doi.org/10.1038/s41598-022-20198-4


86. Z. Wu, Social distancing is a social dilemma game played by every individual against his/her population, PLoS ONE 16: e0255543, 2021, https://doi.org/10.1371/journal.pone.0255543


99. C. Kadelka, Projecting social contact matrices to populations stratified by binary attributes with known homophily, Mathematical Biosciences and Engineering 20: 3282-3300, 2023, https://doi.org/10.3934/mbe.2023154

Equilibrium Strategies w/o Social Distancing

- $\delta_i = 1$
- $\delta_i = 0$
- $\delta_i = 0.5$

Active Time (Hours)

Social Activities

Figure
Small World Network: $m = 200, K = 6, b = 0.3$

Distribution of Degrees of Nodes in Network

Figure
Figure
Figure