Social norm dynamics in a behavioral epidemic model

Anonymous Author(s)
Submission Id: «submission id»

ABSTRACT

Understanding the social determinants of preventive behavior is vital for epidemic modeling and effective policy making. Traditional models emphasize imitation or rational trade-offs, but recent evidence highlights the role of social norms. We develop a behavioral epidemic model of seasonal disease on multilayer networks, where vaccination decisions combine Experience Weighted Attractor (EWA) learning with evolving social norms. The framework distinguishes descriptive norms (what others do) from injunctive norms (what others think ought to be done), while incorporating cognitive dissonance, social projection, and memory. Simulations show that norm dynamics yield markedly different vaccination uptake and infection levels compared to payoff-driven learning. Injunctive norms exert stronger and more persistent effects than descriptive norms, sustaining vaccination even at low prevalence. Interventions targeting injunctive expectations improve outcomes, while those on descriptive norms may be weaker or counterproductive. Norm-based models, once empirically validated, can better capture human behavior and guide strategies for collective action problems beyond pandemics.

KEYWORDS

Agent-based modeling, Behavioral epidemic modeling, Social norm dynamics, Experience-Weighted Attraction learning, Vaccination behavior, Multilayer networks, Collective action problems

1 INTRODUCTION

In 2019 the World Health Organization identified vaccine hesitancy as a major global health threat [84], a concern intensified by COVID-19 [79] and recent influenza outbreaks. Like combating misinformation or climate change, vaccination is a collective-action problem under uncertainty. Formal rules can help, but are often too slow, costly, or unenforceable at scale [80]. This has shifted attention to behavior-changing, participatory approaches—especially social norms—as rapidly adaptive tools to overcome free-riding [53, 58].

Coupled disease–behavior models on multilayer networks capture feedbacks between transmission and protective actions [81, 82]. Early studies examined co-spreading of epidemics and awareness [33, 34]; later work added information valence [36, 83], opinion dynamics [15], and local risk-driven vaccination on coevolving networks [57]. While powerful in modelling the spreading mechanisms, such models are less suited when it comes to designing behavior-targeted interventions because they omit the internal psychological processes behind preventive choices.

Game-theoretic approaches address strategic responses—vaccine uptake, distancing, bounded rationality—under changing prevalence [5, 64], including cognitive biases [77] and social imitation [51, 70]. These frameworks explain phenomena like free-riding [6] but typically operationalize “social influence” narrowly (e.g., imitation), underrepresenting social-psychological mechanisms—cognitive dissonance, social projection, and second-order expectations—that shape adherence to peers and authorities. Cognitive dissonance can shift both behavior and attitudes [23]; people infer others’ intentions via theory of mind [2] and project from self to others [45]. Realistic decision models should therefore track beliefs about peers’ actions and expectations [7, 14, 75]. We follow the standard distinction between descriptive norms (what others do) and injunctive norms (what others think ought to be done) [14]. It is known that their alignment enhances effectiveness of vaccination uptake nudges [7, 56].

Evidence links pro-vaccination norms to intentions and uptake across multiple diseases—influenza [25, 32, 54, 59], HPV [17], and COVID-19 [1, 35, 52, 71]—and public-health guidance increasingly recommends norm-based strategies [9, 69]. In modeling work, descriptive norms often appear as social learning [4, 19], whereas injunctive norms—rooted in normative expectations—have seen limited treatment due to modeling challenges [3, 48, 55]. Memory effects matter in decision models [18] as well, but are rarely combined with norm dynamics; exceptions exist [63], though recent psychologically grounded formulations offer more realistic microfoundations [27].

A growing literature develops cognitively plausible, empirically calibrated Agent-Based Models (ABM) to study norm emergence, downward causation/ “immergence,” and long-run intervention effects [8, 11, 12, 47, 62, 74, 88]. In a recent study [85], Woike et al. showed with their ‘transmission game’ that interventions communicating injunctive social norms substantially reduced risk-taking, while descriptive norm information and case counts were ineffective or even counterproductive. Building on this literature, we construct an ABM that couples a behavioral epidemic model with realistic social-norm dynamics [27], in the spirit of [62], to study vaccination intention and uptake across seasonal outbreaks.

We first present the model—decision process, EWA learning, and norm dynamics—then analyze how norm dynamics shape infection rates and vaccination coverage, including scenarios with stubborn adherence and external interventions. The framework is amenable to lab validation [13]. Our contributions are:

(1) *Co-evolving disease–norm coupling with state-dependent weights:* Following [39], we let the weights on learning, personal norms, and descriptive/injunctive norms adapt to perceived safety and local stability, rather than remain fixed [82]. We link empirical vs. normative emphasis to risk and neighbor stability [31, 38], and extend these dependencies to norm-update rates as suggested in [30], potentially explaining asymmetric norm shifts [78].
(2) *Psychologically grounded norm dynamics in a behavioral epidemic model:* We jointly model descriptive and injunctive norm dynamics via cognitive dissonance, social projection,

Proc. of the Adaptive and Learning Agents Workshop (ALA 2026), Aydeniz, Delgrange, Mohammedalamen, Yang (eds.), May 25 – 26, 2026, Paphos, Cyprus, https://alaworkshop2026.github.io/. 2026.

and logical consistency constraints, focusing on internal decision mechanisms rather than parallel information/opinion diffusion [21, 33]. Norm–risk coevolution alters long-run outbreak size and calls for experimental tests vis-à-vis designs with exogenous risk [73].

(3) *Relative importance of norm types over repeated decisions:* With seasonal repetition and learning, injunctive norms have a stronger long-run impact on outbreak reduction than descriptive norms, aligning with evidence that normative expectations can more powerfully drive cooperation [73], and qualifying results observed under one-shot/high-risk settings [31, 38].

(4) *External influence via nudging on norms:* In line with [29, 76], we model interventions that act on personal and social norms rather than information flows [81, 82], with implications for related domains like misinformation [50].

2 MODEL

During a pandemic, two distinct dynamics are taking place, the epidemic spreading and the related human behavioral response, which are inherently coupled. A common approach to studying the spread of a disease is as a dynamical system on a network of individuals, the network of physical contacts. For the spreading process, the simple Susceptible-Infected-Recovered (SIR) model is often assumed, which is the one we adopt as well, and will be described in the next paragraph. The behavioral change dynamics are also usually described as a dynamical system on a network, termed the social interactions network. In this work, while we have not undertaken a detailed study of the role of the network structure on the dynamics, we considered realistic synthetic networks as you can see in Fig. 1. For the physical layer we assumed a small world topology, motivated by the work in [67] and we assumed a mean degree of 6 and rewiring probability of 0.3. For the social layer we assumed the network to be a Klimek-Thurner network, motivated by [44], where we assumed the same parameter values considered in this work. We modified the social layer's network such that it would have a significant overlap with the small-world network in the physical layer, to model the fact that many physical contacts correspond also to information exchange. We also examined other topologies, such as that of Erdos-Renyi random networks, without changing significantly the results obtained.

We assume that at the beginning of each season individuals decide whether to vaccinate or not. One individual is initially infected. We let the setup run for a number of seasons until we observe the system to reach equilibrium. For each season, we run several simulations of the SIR model, using the event-driven algorithm in [43, 51], in order to obtain the relevant probabilities of infection for each agent. The SIR model is the canonical model of diseases that makes people immune upon recovery. Parameter β represents the transmission rate, while μ is the recovery rate. The latter only sets the time scale of the outbreak, and since we are only interested in the final infected fraction, we set $\mu = 1$ w.l.g.

To study the behavioral change dynamics we employ a game theoretical model, given that the vaccination strategy of each individual as well as the corresponding payoffs depend on other individuals decisions. What is more, the decision of getting vaccinated is usually based not only on the awareness of the transmission severity

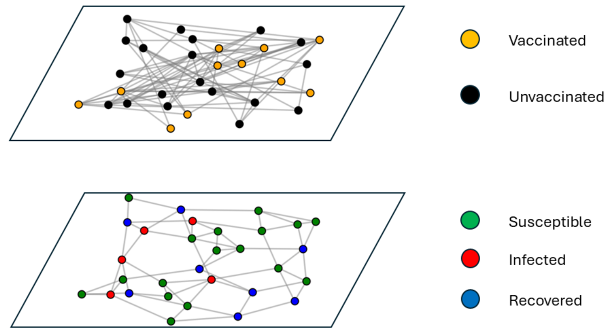

Figure 1: Dynamics of the vaccination game on a two-layer network. In the first layer, the "physical" layer, the seasonal epidemic spreads. These dynamics are governed by a Susceptible-Infected-Recovered (SIR) compartmental model. In the second layer, the "vaccination" layer, the agents receive information from the first layer, as well as from their own layer and they decide whether to vaccinate or not. We assume a small world network degree distribution for the physical layer and a Klimek-Thurner degree distribution for the network of social interactions, which according to the experimental literature are both realistic synthetic networks for the corresponding processes. We consider that the layer of social interactions includes a significant part of the links in the physical layer indicating that many of the physical contacts are also channels of information.

of an epidemic alone, but also on personal beliefs and opinions, that sometimes can be affected by others, the social norms. In the decision-making model that we develop here, we take into account the role of the social norms, and we model their dynamics following the work in [27]. In this work, each individual i is characterized by i) an attitude y_i which gives his personal belief about the most appropriate intention in a given social situation, ii) a belief (an expectation) $\widetilde{x_i}$ about the average intentions of peers (empirical expectations), iii) as well as a second order belief $\widetilde{y_i}$ about the average attitude of their peers (normative expectations).

As was previously stated, the focus of this work is on seasonal diseases. After running several times the SIR dynamics on the physical network, as explained above, we are in a position to obtain descriptive statistics about the epidemiological state of each agent at the end of each season. Then combining this information with the social norms of the agent, the decision making part regarding the vaccination question takes place for the next season. While, if we were aiming to study the role of the dynamics of social norms in the predictive capacities of a model, social norms should be continuously updated in order to capture exogenous events that could alter social norms throughout the year, here our goal is simply to model the effect that the decisions of others have on the updating of the state of the social norms, hence this update is assumed to take place only when decisions are made, i.e. at the beginning of each season. A schematic diagram of the algorithm described in this section is shown in Fig. 2.

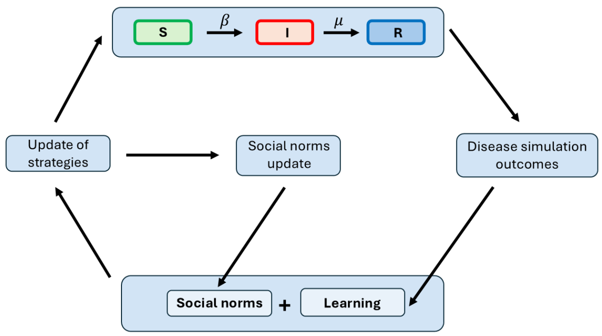

Figure 2: Schematic of the algorithm describing the dynamics of the model. We first run the dynamics several times on the physical layer over a whole infection season. This gives information regarding the probability of being infected during each season. After the end of the season and before the beginning of the next one the agents make a decision regarding vaccination based on the information received from the previous season, by taking into account both material considerations (learning) and normative considerations. Then they also update the norms that they will use at the end of the season based on the decisions taken. Then the next season begins.

The decision making process

Here, we define the dynamics for the total intention $x_i(t)$ of an agent to vaccinate, which should take into account both material considerations as well as norm considerations. $x_i(t)$ corresponds to the probability with which an agent selects to be vaccinated. In [31, 38], the authors noted that, communities under risk, tend to base their decision more on what is (experience) or what they think is actually happening (descriptive norms) rather than considering the role of their beliefs and the beliefs of others (injunctive norms). In the lack of experimental evidence on how this weighting between empirical and morally based considerations might depend on the risk factors, we assume a simple linear relation between the risk perceived and the weight assigned on empirical factors, a usual assumption in norm-utility modelling [76]. Therefore we split the contribution to the intention $x_i(t+1)$ update as follows:

$$x_i(t+1) = (1 - S_i(t))\, x_i^{empirical}(t) + S_i(t)\, x_i^{injunctive}(t) \quad (1)$$

which depends on the safety parameter $S_i(t)$ that quantifies the safety an agent feels and takes values $0 < S_i(t) < 1$. The safety is assumed to be given by the following formula

$$S_i(t) = \begin{cases} 1 - \widehat{I}_i(t) & if\ agent\ unvaccinated \\ 1 - \widehat{I}_{i,neighbors}(t) & if\ agent\ vaccinated \end{cases} \quad (2)$$

where $\widehat{I}_i(t) = \frac{n_{sim}^{inf}(t)}{n_{sim}}$ with n_{sim} the total number of outbreak simulations and $n_{sim}^{inf}(t)$ the number of outbreak simulations when i got infected in the previous cycle, and $\widehat{I}_{i,neighbors}(t) = \frac{n_{sim}^{inf,neigh}(t)}{n_{sim}}$ with $n_{sim}^{inf,neigh}(t)$ the sum of the average number of infected neighbors over all simulation runs.

We define the contribution from what is or what agents think is actually happening, i.e. the empirical contribution to the intention, as

$$x_i^{empirical}(t) = (1 - \phi_i(t))\, p_i^{learn}(t) + \phi_i(t)\, \widetilde{x}_i(t) \quad (3)$$

where we see contributions from the agent's learning of material payoffs $p_i^{learn}(t)$ and what the agent thinks that the rest of the community actually chose on average $\widetilde{x}_i(t)$ (empirical expectations), the dynamics of which are defined in the next subsections. We assume $0 < \widetilde{x}_i(t) < 1$. We weigh these two contributions considering whether the agent will base his decision on his own information or on information obtained from his neighbors, and this is determined by $\phi_i(t)$

$$\phi_i(t) = \sqrt{\phi_i^{change}(t)\, \phi_i^{consensus}(t)} \quad (4)$$

$\phi_i^{change}(t)$ is the change-detector function and is defined as in [39]. This function is related to the surprise function $Q_i(t)$ through $\phi_i^{change}(t) = 1 - \frac{1}{2}Q_i(t)$ where

$$Q_i(t) = \sum_{k=1}^{m_{-1}} \left(h_i^k(t) - q_i^k(t) \right)$$

with

$$q_i^k(t) = I\left(s_{-i}^k, s_{-i}(t)\right)$$

where $I(\cdot)$ symbolizes the indicator function and s_{-i}^k is a vector of dimensions of the number of neighbors of agent i whose each entry contains a unit vector with 0 for all entries except for strategy k. $s_{-i}(t)$ is a vector of dimensions of the number of neighbors of agent i, containing the actual choices at time t of the neighbors. Hence $q_i^k(t)$ is a vector of dimensions of the number of neighbors of agent i, with 1s for all neighbors who had chosen strategy k at time t. On the other hand,

$$h_i^k(t) = \frac{\sum_{\tau=t-m}^{t} I\left(s_{-i}^k, s_{-i}(\tau)\right)}{m}$$

which is a vector that records the historical frequencies of the choices by other players, over a span of period of m steps, which is the size of the memory of the agent. $\phi_i^{consensus}(t) \in [0,1]$ is the consensus function given by

$$\phi_i^{consensus}(t) = |2 * X_i - 1| \quad (5)$$

where X_i is the average fraction of neighbors vaccinated as observed by the focal individual, and hence $\phi_i^{consensus}(t)$ is equal to 1 when neighbors agree and 0 if half are vaccinated and the other half not.

The intuition behind $\phi_i(t)$ is the following: when the surprise function is high but the change-detector function is low—meaning that the focal agent's neighbors frequently switch their vaccination choices—the agent should rely more on their own judgment rather than on social cues, since the neighbors are not in a stable state, possibly due to strong environmental fluctuations. Similarly if there is great heterogeneity in the neighbors actions then the focal agent should also resort to his own judgement. To capture this effect, we use the geometric mean, which ensures that both stability and consensus among neighbors must be present for their behavior to be a reliable source of information.

The contribution from the belief dependent part of the agents, which is the mechanism activated when the agent is not at risk takes the following form

$$x_i^{injunctive}(t) = (1-\phi_i(t))\, y_i(t) + \phi_i(t)\, \widetilde{y}_i(t) \quad (6)$$

where again the idea is to split the contributions between the effect of own beliefs and the beliefs of others using $\phi_i(t)$. We assume $0 < \widetilde{y}_i(t), y_i(t) < 1$. The dynamics of the norm variables, $y_i(t)$ and $\widetilde{y}_i(t)$ are also defined in the next subsections. A diagram with the contributions to the total intention is shown in Fig. 3.

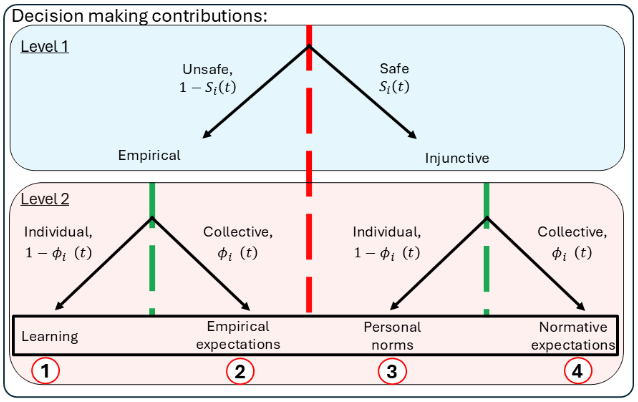

Figure 3: Schematic diagram of the decision making process. The contributions to the final intention of vaccinating from both the learning process and the social norms is summarized with the relevant weight of each one of them. The main assumptions that determine the contribution of each factor are a) The more unsafe an agent feels (i.e. the lower $S_i(t)$ is, where $S_i(t)$ depends on the probability of an agent getting infected in the previous season), the more probable is that he will base his decision on empirical factors, i.e. through learning or through his empirical expectation, rather than normative factors, such as his personal believes or his normative expectation. b) The more heterogeneous and unstable his environment is (i.e. the more the decisions of agents differ among themselves and also differ in the immediate previous round from their average choices in the few previous rounds), the more probable is for an agent to base his decision on individual factors, such as learning (from his own experience) and his personal believes, rather than resorting to what other agents are doing.

We should note that here we consider one of the simplest mechanisms to model the decision of an individual to trust individual or collective factors, namely the surprise dependent mechanism. In this case, there is an implicit assumption that, whenever an individual perceives that his neighbours are changing their behaviour, then he trusts that he can individually make the correct decision in this changing environment, by either learning from his personal experience, or trusting his personal norm. By considering this simple surprise dependent mechanism however, we omit other potential influences that could guide an agent's decision to be more individualistic or more collectivistic, such as for example the trust that he has on his neighbors' abilities to make the correct decision in a changing epidemiological situation. However this would depend on a number of other factors, such as for example how deadly the disease is and it would make the model more complicated.

Numerous social and behavioral science theories offer insights into behavior and change [16]. Our approach, building on earlier Continuous Opinions and Discrete Actions (CODA) models [49], links individual actions to attitudes influenced by observed behaviors, aligned with social psychology [41], which prioritize psychological rules over perfect rationality in adoption decisions. This perspective was also applied in [76]. Prior mathematical models often ignore the dynamic nature of attitudes, leading to inaccuracies in adoption predictions and policy efficacy.

The experience dependent mechanism

Human social behavior is influenced by model-free and model-based reinforcement learning [10], combining both past experiences and future goals. These can both be captured by a learning mechanism known as the Experience-Weighted-Attractor (EWA) model, as long as certain conditions, stated in [26] are fullifilled. EWA is a learning mechanism that encapsulates both reinforcement learning [72] and belief learning [22], which is to say that actions are reinforced based on both how well they actually performed, i.e. based on their actual material payoffs, but also based on what their potential outcomes could have been had they been chosen, i.e. their forgone material payoffs. Here, we assume an equal weight for actual and forgone material payoffs. Agents estimate forgone material payoffs based on the average number of infected neighbors, without access to neighbors' payoffs in agreement with experimental findings [68]. We further incorporate memory and bounded rationality into EWA. Agents evaluate average payoffs over the last m decision cycles, with more recent outcomes weighted more heavily when risk is high (safety-dependent memory decay). Finally, rationality is modulated by perceived safety: in high-risk states, agents are less able to discriminate between payoffs (captured by a sensitivity parameter κ).

The payoffs are therefore estimated as following:

$$\Pi_i^{Unvac}(t) = \begin{cases} 1 - c_I \widehat{I}_{i,neighbors}(t) & if\ vaccinaed \\ 1 - c_I \widehat{I}_i(t) & if\ not\ vaccinated \end{cases} \quad (7)$$

$$\Pi_i^{Vac}(t) = 1 - c_V$$

and based on this, the materially-motivated intention of the agent i at time t to vaccinate or not, is assumed to be given by the commonly employed Quantal Response Equilibrium (QRE) for binary choices, which is simply given by the logistic (softmax) function

$$p_i^{learn}(t) = \frac{1}{1 + e^{-\frac{\pi_i^{Unvac}(t) - \pi_i^{Vac}(t)}{\kappa}}} \quad (8)$$

Here $\pi_i^{Vac}(t)$ is the average payoff for the vaccinating option over the last m rounds, while $\pi_i^{Unvac}(t)$ is the average payoff received for not vaccinating over the last m rounds $\pi_i^{Unvac}(t)$, given by

$$\pi_i^{Unvac}(t) = \sum_{j=0}^{m} \frac{(S_i(t))^j}{\sum_{n=0}^{m} (S_i(t))^n} \Pi_i^{Unvac}(t-j) \quad (9)$$

where the safety parameter $S_i(t)$ defined above plays the role of a memory decaying function. The idea behind the use of the QRE formula, is that when $p_i^{learn}(t) = \frac{1}{2}$, this represents no preference towards either vaccinating or not vaccinating. Then whenever

$\pi_i^{Unvac}(t) > \pi_i^{Vac}(t)$ or $\pi_i^{Unvac}(t) < \pi_i^{Vac}(t)$ holds, the preference should shift towards not vaccinating or vaccinating respectively.

Social norm dynamics

To integrate social norm dynamics into the EWA model, first of all, we separate material payoffs from social norm dynamics. Experiments show that people represent the preferences and beliefs of others, i.e. their norm variables, separately from their own [37]. A recent work directly measures these variables in behavioral experiments [73]. In line with experimental evidence, we treat personal attitudes, empirical expectations (beliefs about others' actions), and normative expectations (beliefs about others' attitudes) as separate state variables rather than collapsing them into a single "imitation" process. We are interested in the dynamics of these variables:

$$
\begin{aligned}
y_i' &= y_i + \left[\underbrace{C_i^{11}(x_i - y_i)}_{cognitive\,dissonance} + \underbrace{C_i^{12}(X_i - y_i)}_{conformity\,w/peers} + \underbrace{C_i^{13}(G_i^1 - y_i)}_{conformity\,w/auth.} \right] \\
\widetilde{y}_i' &= \widetilde{y}_i + \left[\underbrace{C_i^{21}(y_i - \widetilde{y}_i)}_{social\,projection} + \underbrace{C_i^{22}(X_i - \widetilde{y}_i)}_{learning\,about\,others} + \underbrace{C_i^{23}(G_i^2 - \widetilde{y}_i)}_{conformity\,w/auth.} \right] \\
\widetilde{x}_i' &= \widetilde{x}_i + \left[\underbrace{C_i^{31}(\widetilde{y}_i - \widetilde{x}_i)}_{logic\,constraints} + \underbrace{C_i^{32}(X_i - \widetilde{x}_i)}_{learning\,about\,others} + \underbrace{C_i^{33}(G_i^3 - \widetilde{x}_i)}_{conformity\,w/auth.} \right]
\end{aligned}
\tag{10}
$$

where the prime signifies the next time step, X_i as above and C_i^{lj} represents non-negative individual-specific constant coefficients. The dynamics follow a de-Groot–type update rule where each variable is pulled toward three psychological forces: cognitive dissonance, social projection, and conformity. For example, attitudes shift not only to align with personal intentions but also to justify past actions, while expectations adjust both to peers' behavior and to inferred attitudes of others. Here the 'cognitive dissonance' term acts to reduce the mismatch of the ego's actions and their beliefs about themselves. The 'social projection' term captures the ego's belief that others are probably similar to themselves [45]. The 'logic constraints' term reduces a mismatch between the ego's beliefs about actions and beliefs of others [24]. The 'conformity w/ peers' and two 'learning about others' terms move the corresponding attitude and beliefs closer to the observed average behavior X_i among peers [42]. The 'conformity w/ authority' terms move the corresponding attitudes and beliefs closer to the promoted 'standard' G_i^l. Note that cognitive dissonance brings individuals' intention x_i closer to their attitude y_i and simultaneously changes their attitude y_i to justify the action previously chosen [60].

$$
\begin{aligned}
y_i' &= y_i + \xi_i^1 \left[\widehat{C_i^{11}} x_i + \widehat{C_i^{12}} X_i + \widehat{C_i^{13}} G_i^1 - y_i \right] \\
\widetilde{y}_i' &= \widetilde{y}_i + \xi_i^2 \left[\widehat{C_i^{21}} y_i + \widehat{C_i^{22}} X_i + \widehat{C_i^{23}} G_i^2 - \widetilde{y}_i \right] \\
\widetilde{x}_i' &= \widetilde{x}_i + \xi_i^3 \left[\widehat{C_i^{31}} \widetilde{y}_i + \widehat{C_i^{32}} X_i + \widehat{C_i^{33}} G_i^3 - \widetilde{x}_i \right]
\end{aligned}
\tag{11}
$$

where $\xi_i^l = \sum_j C_i^{lj}$ is interpreted as the rate at which the l^{th} variable is updated, and $\widehat{C_i^{lj}} = \frac{C_i^{lj}}{\sum_k C_i^{lk}}$ such that $\sum_j \widehat{C_i^{lj}} = 1$. In the rest of the article we assume that $\xi_i^1 = 0.01, \xi_i^2 = 0.1$ and $\xi_i^3 = 1$, to model the fact that personal norms evolve slower than the rest while empirical expectations are the ones that evolve the fastest, in agreement with intuition. Results hold for a large range of these parameters as long as the relation between them holds $\xi_i^3 > \xi_i^2 > \xi_i^1$. Rewriting the equations in this form, we see that the coefficients $\widehat{C_i^{lj}}$ can be interpreted as the weights of the effect that each of the three variables in the equations can have on the evolution of the l^{th} variable. To adjust the dynamics to the problem that we have at hand, i.e. to the dynamics of social norms of a group of individuals under the collective risk of an epidemic whose dynamics are governed as explained above, we define the coefficients $\widehat{C_i^{lj}}$ in the way detailed below.

First, for the time being, we ignore the role of any external authorities. Hence we set all $\widehat{C_i^{l3}} = 0$, which subsequently implies that $\widehat{C_i^{l2}} = 1 - \widehat{C_i^{l1}}$. Then, we note that for each variable, its evolution depends, besides its previous value, on two other variables, which can be divided into two categories. The variables depending on individual i's social norms i.e. x_i, y_i, $\widetilde{y}_i$, and the variables that depend on the action of the neighbors X_i . Hence, we interpret $\widehat{C}_{i1}$ as the part of the updating of the i^{th} variable attributed to looking inward, at his own previous norms, and subsequently, $1 - \widehat{C_i^{l1}}$ is the weight that the agent puts on the neighbors influence. Alternatively, based on the original formulation of the dynamic equations for the social norms, we can interpret $\widehat{C}_{i1}$ and $\widehat{C}_{i2}$ as the weights that the agent puts on the psychological processes involving only his norms, and the psychological processes involving what he observes in his surroundings respectively. Essentially this amount to assigning weights on inward vs. outward orientation. Reinterpreting the coefficients in this way allows us to relate them to quantities obtained by the physical spreading process taking place, rather than assuming some arbitrary value for them. We assume that

$$C_i^{l1} = 1 - \phi_i(t) \tag{12}$$

and hence $C_i^{l2} = \phi_i(t)$ where $\phi_i(t)$ as defined in the previous section. We link these weights to the change-detector function and the consensus function, to model the fact that in unstable environments agents rely more on personal anchors, whereas in stable contexts they put more weight on their peers. This operationalizes the psychological intuition that norm updating is faster when social signals are clear and consistent, and slower when the environment is noisy. A diagram of the contributions to the dynamics of the social norms can be found in Fig. 4.

3 RESULTS

We run 1000 simulations of the Susceptible-Infected-Recovered (SIR) model on a network of $N = 500$ nodes, using the event-driven algorithm in [43, 51], for each season. We assume the rationality parameter to be $\kappa = 0.1$ and the vaccination cost $c_V = 1$ and $c_I = 0.1$ in all of the simulations that follow. Whenever not stated, memory length is assumed to be $m = 4$ and $\beta = 6$. The initial values of the three norms associated with each agent are drawn independently from a uniform distribution. Finally, in all the cases studied, we made sure that the system was simulated for sufficiently many seasons such that evolutionary equilibrium was reached, in particular that the percentage of vaccinated individuals did not change by more than 0.01 in the last 50 seasons and with a maximum

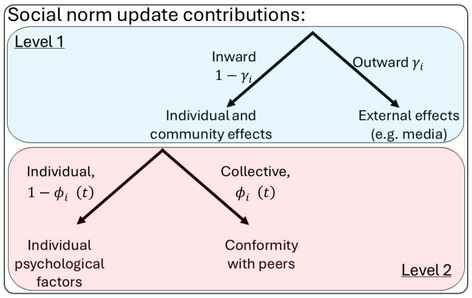

Figure 4: Schematic diagram of the contributions to the dynamics of the social norm variables. The contributions of each one of the factors that drive the dynamics of the social norms is summarized. The main assumptions that determine the contribution of each factor are a) each agent is characterized by some tendency to resort to external factors to update the values for her norms, expressed by some probability γ_i, while alternatively she considers factors related to her individual factors (i.e. her personal believes and social norms) or her immediate peers' actions b) the more unstable her environment is (i.e. the more the decisions of agents in the immediate previous round differs from their average choices in the few previous rounds), the more probable is for an agent to base her decision on individual factors rather than conforming to what other agents are doing.

of 200 seasons. In all of the figures presented in this section we include the lower and upper quartile range as well.

The role of social norms. Initially, we study the role of memory on epidemic outcomes without considering social norms. Figure 5A shows that the infected fraction increases with the infectivity rate β, remaining close to zero for small β and saturating at high values. Longer memory reduces the outbreak size, as agents incorporate a richer history of payoffs when deciding whether to vaccinate. Figure 5B displays the corresponding vaccination coverage: for small β, coverage is close to zero in the absence of norms, while for large β it rises toward 0.5, driven by panic effects. Increasing memory boosts coverage further, though not linearly. This is both due to network effects, but also due to the fact that high β values reduce the safety factor $S_i(t)$ and thus amplify the role of recent outcomes in shaping decisions. When social norm dynamics are introduced, both infection levels and coverage change significantly: norms sustain a non-trivial fraction of vaccination (around 20%) even at low β, reflecting the persistence of initial heterogeneous attitudes. Figure 6 illustrates how the three norm variables evolve. As expected, since we are looking at the equilibrium values, the three norms that are freely evolving coincide between them, as well as with the vaccination intention.

Zealots. Figures 7 and 8 investigate the effect of zealotry—agents being stubborn with respect to one social norm variable—on epidemic outcomes. Fig. 7A shows that the outbreak size is smallest when all three norms coevolve dynamically, while fixing one norm increases the infected fraction. The effect is strongest when zealotry applies to normative expectations ($\tilde{y}$), which produce the highest infection levels, whereas zealotry in empirical expectations ($\tilde{x}$) yields

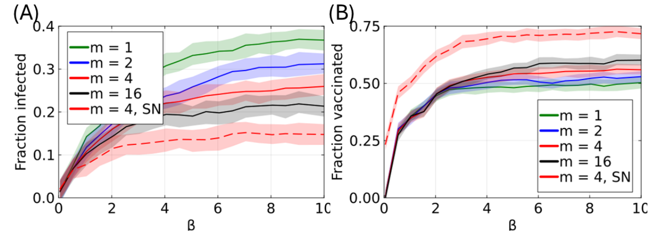

Figure 5: Role of memory and social norms on the infected fraction and vaccination coverage. (A) Infected fraction as a function of infectivity rate β. As β increases, the fraction infected grows and eventually plateaus, while remaining near zero for small β (no clear phase transition). Larger memory reduces outbreak size, and for fixed memory ($m = 4$), adding social norm dynamics further decreases infections. (B) Vaccination coverage as a function of β. For large β, coverage stabilizes around 0.5 due to panic effects. With social norms, even small β values sustain $\sim 20\%$ coverage, since initial norm values are uniformly random and persist through their dynamics.

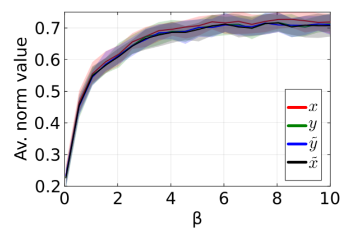

Figure 6: Average values of the social norm variables y, $\widetilde{y}$ and $\widetilde{x}$ as well as the vaccination intention x over all agents. All three of the social norms, as well as the vaccination intention x coincide in the long run as expected.

the smallest increase. Fig. 7B further illustrates that vaccination coverage reflects these dynamics: when personal beliefs (y) or normative expectations ($\tilde{y}$) are fixed, coverage converges to 0.5—the mean of the uniformly distributed fixed norms—even for very small values of the infection rate β. By contrast, zealotry in empirical expectations has little influence on coverage, which remains closer to the adaptive case. Fig. 8 examines how zealotry affects the evolution of the norms themselves. Fixing empirical expectations ($\tilde{x}$) leaves personal and normative norms largely unaffected, while zealotry in personal beliefs (y) exerts stronger influence on the others. Strikingly, zealotry in normative expectations ($\tilde{y}$) dominates the system, driving both personal beliefs and empirical expectations toward its fixed value. These results align with experimental evidence that normative expectations play a pivotal role in norm emergence and coordination, while empirical expectations are less influential [7, 85].

The role of external factors. We study the role of an external factor affecting the evolution of social norms, assuming that the external effect can be present independently of the epidemiological state of the community. This is introduced in the following way

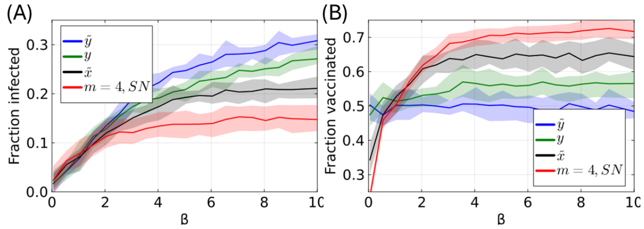

Figure 7: Role of zealotry in social norms. (A) Infected fraction as a function of infectivity rate β when agents are zealots with respect to y, $\widetilde{y}$, or $\widetilde{x}$. In all cases the outbreak size exceeds that with full dynamics, with the strongest increase from zealotry in normative expectations $\widetilde{y}$ and the weakest from zealotry in empirical expectations $\widetilde{x}$. (B) Vaccination coverage as a function of β under zealotry. Zealotry in $\widetilde{y}$ drives coverage to 0.5, the mean fixed-norm value, even for very small β.

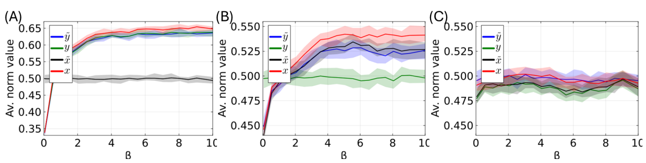

Figure 8: Average of social norm variables y, $\widetilde{y}$ and $\widetilde{x}$ over all agents, as a function of the infectivity rate β for the cases of zealotry considered on each one of the social norm variables. (A), zealotry on the empirical expectation $\widetilde{x}$, (B) zealotry in the personal norms y, and (C) zealotry in the normative expectations $\widetilde{y}$. We see that the empirical expectations $\widetilde{x}$ have the least effect on the other social norms, while the normative expectations $\widetilde{y}$ the most, practically driving the other variables.

$$
\begin{aligned}
y_i' &= y_i + \xi_i^1 \left[\gamma_i^1 G_i^1 + \widetilde{C_i^{11}} x_i + \widetilde{C_i^{12}} X_i - y_i \right] \\
\widetilde{y}_i' &= \widetilde{y}_i + \xi_i^2 \left[\gamma_i^2 G_i^2 + \widetilde{C_i^{21}} y_i + \widetilde{C_i^{22}} X_i - \widetilde{y}_i \right] \\
\widetilde{x}_i' &= \widetilde{x}_i + \xi_i^3 \left[\gamma_i^3 G_i^3 + \widetilde{C_i^{31}} \widetilde{y}_i + \widetilde{C_i^{32}} X_i - \widetilde{x}_i \right]
\end{aligned}
\tag{13}
$$

where $0 \le \gamma_i^l \le 1$ for $l \in \{1,2,3\}$ is the strength of the external factor signal, and G_i^l is the target value for the norms. $\widetilde{C_i^{lj}} = \left(1-\gamma_i^l\right)\widehat{C_i^{lj}}$, where $\widehat{C_i^{l1}}$ and $\widehat{C_i^{l2}} = 1 - \widehat{C_i^{l1}}$ as before. Note that the condition that the coefficients of the variables sum up to 1 is still satisfied such that these can be still interpreted as relative contributions. In our studies, we assume the same strength γ_i^l and target G_i^l for all agents, i.e. $\gamma_i^l = \gamma^l$ and $G_i^l = G^l$ $\forall i \in \{1, ..., N\}$.

Figures 9–11 analyze how external influences on social norms alter epidemic outcomes. Fig. 9 shows that when an external factor plays down disease severity by shifting norms toward a lower target ($G = 0.6$), outbreaks become larger if the intervention acts through personal norms or injunctive expectations, while effects through empirical expectations remain negligible. Moving to Fig. 10, we observe richer dynamics. Panel A demonstrates that the outbreak size depends not only on which norm is targeted but also on the strength of the intervention (γ) and its target value (G). Driving personal norms or injunctive expectations toward lower values reliably increases outbreak size, whereas targeting empirical expectations can sometimes backfire against the external factor: for intermediate values of G, the community resists, and infection levels fall below the baseline without external influence. Panel B further highlights these nonlinearities, showing that for sufficiently high target values ($G \approx 0.7$) outbreaks shrink under all intervention types, with empirical expectations providing the strongest buffer, while for low G values the outbreak size grows, especially when personal or injunctive norms are targeted. The nonlinear responses in Fig. 10 illustrate a boomerang effect [46, 65, 86]: external attempts to weaken vaccination norms sometimes backfire, leading to stronger compliance and reduced infection, akin to backfire effects reported in descriptive norm interventions [61]. Finally, Fig. 11 examines how the average values of each norm evolve under external driving. Only the targeted norm consistently converges to the imposed value, but interventions on injunctive expectations also drag the other norms along, underscoring their coordinating role.

Taken together, these results reveal that external factors do not affect all norms equally: interventions on injunctive norms spread their influence across the system, while empirical expectations can generate resilience, even neutralizing attempts to downplay risk. This is in agreement with the experimental results in [73] and the subsequent theoretical analysis in [66] where the authors found that the normative variables play a key role in the emergence of a norm.

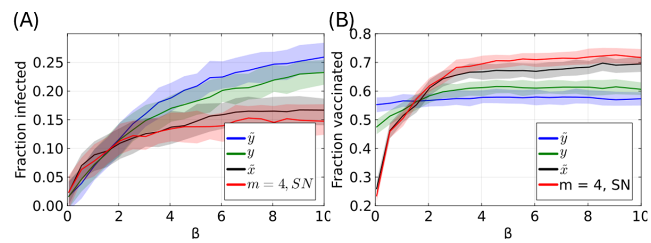

Figure 9: Role of external factors on social norms. (A) Infected fraction as a function of the infectivity rate β for three distinct cases corresponding to the cases where an external factor affects each one of the social norm variables y_i, $\widetilde{y}_i$ and $\widetilde{x}_i$. We assume that the role of the external factor is to play down the severity of the disease spreading. We assume that the external factor is driving the social norm to the value $G_i = 0.6$, slightly lower than the equilibrium value 0.7 for the social norms observed in Fig. 6. We see that the external factor is successful in increasing the infected fraction when is applied to either the personal norms y_i or to the normative expectations $\widetilde{y}_i$. However when applied to the empirical expectation $\widetilde{x}_i$ the external factor's effect is negligible. (B) Vaccination coverage as a function of the infectivity rate β for the aforementioned cases. The parameters used in the simulations were: $\gamma_i = 0.5$, $G_i = 0.6$.

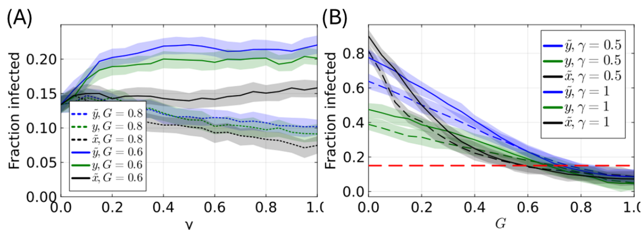

Figure 10: Role of external factor's strength γ and external factor's target G. (A) Outbreak size as a function of the external factor's coupling strength γ, when the intervention acts on one of the three social norm variables: personal norms y_i, normative expectations $\widetilde{y}_i$, or empirical expectations $\widetilde{x}i$. We consider two target values, $G = 0.6$ and $G = 0.8$, corresponding respectively to driving norms toward lower or higher severity perceptions than the baseline case without external influence. When the external factor pushes either personal norms or normative expectations toward $G = 0.6$, the infected fraction increases. In all other cases, the population reacts adaptively: for sufficiently high coupling strength, the resulting outbreak size becomes smaller than in the absence of external influence. Of particular interest is the effect of driving the empirical expectation $\widetilde{x}$ towards $G = 0.6$ (lower than the no external influence case) which does not achieve to increase the outbreak size. (B) Infected fraction as a function of the target value G for the three intervention scenarios, with coupling strengths $\gamma = 0.5$ and $\gamma = 1$. The red dashed line indicates the baseline infection level without external intervention. For large target values (nudging towards vaccination), the infected fraction approaches zero, as expected. For small target values (nudging individuals away from vaccination), outbreak sizes are substantially larger. For intermediate values of G the intervention is more effective through $\widetilde{y}$ while for sufficiently small G, targeting empirical expectations $\widetilde{x}_i$ becomes more effective. When external driving is applied to personal norms or normative expectations, even a modest reduction of the target below the baseline vaccination coverage ($G \approx 0.7$) leads to increased outbreak size. However, when the intervention targets empirical expectations, the community can partially resist: for a range of intermediate target values ($G \approx 0.6 - 0.8$), infection levels are lower than in the baseline case, indicating that descriptive norms can buffer against external attempts to undermine vaccination.

4 CONCLUSION

Epidemic–behavior models often rely on simple imitation or payoff-based reinforcement, which capture certain dynamics but fail to offer a more realistic description of how attitudes, expectations, and social pressures shape preventive decisions, which opens the way for behavioral intervention proposals. To address this gap, we developed a behavioral epidemic model that couples disease spread with a decision-making process grounded in the Experience-Weighted Attraction (EWA) learning framework [10]. EWA combines reinforcement and belief learning by accounting for both realized and foregone payoffs, something that better describes human decision making in experimental settings. We extended this by integrating social norm dynamics, motivated by the distinction between descriptive and injunctive norms in social psychology [7, 27, 56]. The resulting model captures mechanisms such as cognitive dissonance, social projection, and consistency constraints, which explain how individuals align with peers or authorities.

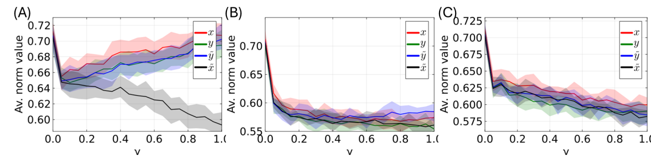

Figure 11: Social norm variables dependence on the external factor's coupling strength γ for the scenarios of coupling to one of the three social norm variables y, $\widetilde{y}$ and $\widetilde{x}$ each time. We assume a $G = 0.6$ for each case. The average of the social norm variables over all agents is represented. (A) Here the external driving is on the empirical expectation $\widetilde{x}$. As we see its value goes to the target's value when the coupling strength is sufficiently high. the rest of the social norms do not follow it. (B) Here the external driving is on the normative expectation $\widetilde{y}$. The other social norms follow its value. (C) Here the external driving is on the personal norms y. The other social norms follow its value.

Our analysis shows that considering the coevolving norm dynamics substantially alters the final outbreak size. Furthermore, we find that injunctive norms stabilize vaccination in repeated decision contexts, while descriptive norms mainly track short-term fluctuations [55, 73]. Simulations with zealots demonstrate that all norms reduce infection, though normative expectations dominate while empirical expectations matter least [50, 76, 79, 85]. External interventions targeting personal or normative expectations strongly influence vaccination coverage, whereas descriptive-norm interventions may be weaker or counterproductive. These findings align with experimental evidence, such as the transmission game, where injunctive-norm nudges proved more effective in reducing collective risk-taking [85].

Our framework is not designed for predictive forecasting but for explanatory insight into how social norms coevolve with epidemic risk and shape vaccination behavior. Such insights can inform communication strategies that emphasize stable normative expectations rather than volatile descriptive signals [9, 52, 69]. More broadly, norm-based models—once validated experimentally—can guide interventions for other collective-action challenges, from misinformation to environmental cooperation. Future extensions include adaptive networks [82], multi-layer information and fear propagation [20, 87], homophily in intentions or vaccination status [40], and the role of peer punishment and rewards in norm evolution [28].

REFERENCES

[1] M. Agranov, M. Elliott, and P. Ortoleva. 2021. The importance of Social Norms against Strategic Effects: The case of Covid-19 vaccine uptake. *Econ Lett* 206 (Sep 2021), 109979. https://doi.org/10.1016/j.econlet.2021.109979

[2] Ian Apperly. 2010. *Mindreaders: the cognitive basis of" theory of mind"*. Psychology Press.

[3] Pierpaolo Battigalli and Martin Dufwenberg. 2022. Belief-Dependent Motivations and Psychological Game Theory. *Journal of Economic Literature* 60, 3 (September 2022), 833–82. https://doi.org/10.1257/jel.20201378

[4] Chris T. Bauch and Samit Bhattacharyya. 2012. Evolutionary Game Theory and Social Learning Can Determine How Vaccine Scares Unfold. *PLOS Computational Biology* 8, 4 (04 2012), 1–12. https://doi.org/10.1371/journal.pcbi.1002452

[5] Chris T. Bauch and David J. D. Earn. 2004. Vaccination and the theory of games. *Proceedings of the National Academy of Sciences* 101, 36 (2004), 13391–13394. https://doi.org/10.1073/pnas.0403823101 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.0403823101

[6] Samit Bhattacharyya and Chris T. Bauch. 2011. "Wait and see" vaccinating behaviour during a pandemic: A game theoretic analysis. *Vaccine* 29, 33 (2011), 5519–5525. https://doi.org/10.1016/j.vaccine.2011.05.028

[7] Cristina Bicchieri. 2005. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press. https://doi.org/10.1017/CBO9780511616037

[8] Giangiacomo Bravo, Flaminio Squazzoni, and Riccardo Boero. 2012. Trust and partner selection in social networks: An experimentally grounded model. *Social Networks* 34, 4 (2012), 481–492. https://doi.org/10.1016/j.socnet.2012.03.001

[9] N. T. Brewer, G. B. Chapman, A. J. Rothman, J. Leask, and A. Kempe. 2017. Increasing Vaccination: Putting Psychological Science Into Action. *Psychol Sci Public Interest* 18, 3 (Dec 2017), 149–207. https://doi.org/10.1177/1529100618760521

[10] Colin Camerer and Teck Hua Ho. 1999. Experience-weighted attraction learning in normal form games. *Econometrica* 67, 4 (1999), 827–874. https://doi.org/10.1111/1468-0262.00054

[11] Donald T. Campbell. 1974. 'Downward Causation'in Hierarchically Organised Biological Systems. In *Studies in the philosophy of biology: reduction and related problems*, Francisco Jose Ayala and Theodosius Dobzhansky (Eds.). University of California Press, 179. https://doi.org/10.1007/978-1-349-01892-5_11

[12] Cristiano Castelfranchi. 1998. Simulating with Cognitive Agents: The Importance of Cognitive Emergence. In *mAbs*. https://doi.org/10.1007/10692956_3

[13] G. B. Chapman, M. Li, J. Vietri, Y. Ibuka, D. Thomas, H. Yoon, and A. P. Galvani. 2012. Using game theory to examine incentives in influenza vaccination behavior. *Psychological Science* 23, 9 (2012), 1008–1015. https://doi.org/10.1177/0956797612437606

[14] Robert B Cialdini, Raymond R Reno, and Carl A Kallgren. 1990. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology* 58, 6 (1990), 1015. https://doi.org/10.1037/0022-3514.58.6.1015

[15] Xiangfeng Dai, Peican Zhu, Yangming Guo, and Zhen Wang. 2019. Coevolution of Vaccination Opinions and Awareness Affecting the Spread of Epidemics. *IEEE Access* 7 (2019), 61558–61569. https://doi.org/10.1109/ACCESS.2019.2902856

[16] R. Davis, R. Campbell, Z. Hildon, L. Hobbs, and S. Michie. 2015. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. *Health Psychol Rev* 9, 3 (2015), 323–344. https://doi.org/10.1080/17437199.2014.941722

[17] R. de Visser, L. Waites, C. Parikh, and A. Lawrie. 2011. The importance of social norms for uptake of catch-up human papillomavirus vaccination in young women. *Sex Health* 8, 3 (Sep 2011), 330–337. https://doi.org/10.1071/SH10155

[18] Xiao-Heng Deng, Yi Liu, and Zhi-Gang Chen. 2010. Memory-based evolutionary game on small-world network with tunable heterogeneity. *Physica A: Statistical Mechanics and its Applications* 389, 22 (2010), 5173–5181. https://doi.org/10.1016/j.physa.2010.08.0

[19] Alberto d'Onofrio, Piero Manfredi, and Piero Poletti. 2012. The Interplay of Public Intervention and Private Choices in Determining the Outcome of Vaccination Programmes. *PLOS ONE* 7, 10 (10 2012), 1–10. https://doi.org/10.1371/journal.pone.0045653

[20] Joshua M Epstein. 2014. *Agent_Zero: toward neurocognitive foundations for generative social science*. Princeton University Press.

[21] Fanshu Fang, Jing Ma, and Yanli Li. 2023. The coevolution of the spread of a disease and competing opinions in multiplex networks. *Chaos, Solitons, Fractals* 170 (2023), 113376. https://doi.org/10.1016/j.chaos.2023.113376

[22] Nick Feltovich. 2012. *Belief-Based Learning Models*. 444–447. https://doi.org/10.1007/978-1-4419-1428-6_111

[23] Leon Festinger. 1957. *A theory of cognitive dissonance*. Stanford University Press.

[24] N. E. Friedkin, A. V. Proskurnikov, R. Tempo, and S. E. Parsegov. 2016. Network science on belief system dynamics under logic constraints. *Science* 354, 6310 (Oct 2016), 321–326. https://doi.org/10.1126/science.aag2624

[25] M. M. Galizzi, K. W Lau, M. Miraldo, and K. Hauck. 2022. Bandwagoning, free-riding and heterogeneity in influenza vaccine decisions: An online experiment. *Health Econ* 31, 4 (Apr 2022), 614–646. https://doi.org/10.1002/hec.4467

[26] Tobias Galla and J Doyne Farmer. 2013. Complex dynamics in learning complicated games. *Proceedings of the National Academy of Sciences* 110, 4 (2013), 1232–1236. https://doi.org/10.1073/pnas.1109672110

[27] Sergey Gavrilets. 2021. Coevolution of actions, personal norms and beliefs about others in social dilemmas. *Evolutionary Human Sciences* 3 (2021), e44. https://doi.org/10.1017/ehs.2021.40

[28] Sergey Gavrilets and Peter J. Richerson. 2017. Collective action and the evolution of social norm internalization. *Proceedings of the National Academy of Sciences* 114, 23 (2017), 6068–6073. https://doi.org/10.1073/pnas.1703857114 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1703857114

[29] Sergey Gavrilets and Peter J. Richerson. 2022. Authority matters: propaganda and the coevolution of behaviour and attitudes. *Evolutionary Human Sciences* 4 (2022), e51. https://doi.org/10.1017/ehs.2022.48

[30] Sergey Gavrilets, Denis Tverskoi, and Angel Sánchez. 2024. Modelling social norms: an integration of the norm-utility approach with beliefs dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences* 379, 1897 (2024), 20230027. https://doi.org/10.1098/rstb.2023.0027 arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.2023.0027

[31] Michele J. Gelfand and Jesse R. Harrington. 2015. The Motivational Force of Descriptive Norms: For Whom and When Are Descriptive Norms Most Predictive of Behavior? *Journal of Cross-Cultural Psychology* 46, 10 (2015), 1273–1278. https://doi.org/10.1177/0022022115600796 arXiv:https://doi.org/10.1177/0022022115600796

[32] J. R. Gorman, N. T. Brewer, J. B. Wang, and C. D. Chambers. 2012. Theory-based predictors of influenza vaccination among pregnant women. *Vaccine* 31, 1 (Dec 2012), 213–218. https://doi.org/10.1016/j.vaccine.2012.10.064

[33] Clara Granell, Sergio Gómez, and Alex Arenas. 2013. Dynamical Interplay between Awareness and Epidemic Spreading in Multiplex Networks. *Phys. Rev. Lett.* 111 (Sep 2013), 128701. Issue 12. https://doi.org/10.1103/PhysRevLett.111.128701

[34] Clara Granell, Sergio Gómez, and Alex Arenas. 2014. Competing spreading processes on multiplex networks: Awareness and epidemics. *Phys. Rev. E* 90 (Jul 2014), 012808. Issue 1. https://doi.org/10.1103/PhysRevE.90.012808

[35] S. Graupensperger, D. A. Abdallah, and C. M. Lee. 2021. Social norms and vaccine uptake: College students' COVID vaccination intentions, attitudes, and estimated peer norms and comparisons with influenza vaccine. *Vaccine* 39, 15 (Apr 2021), 2060–2067. https://doi.org/10.1016/j.vaccine.2021.03.018

[36] Dun Han and Xin Wang. 2024. Impact of positive and negative information on epidemic spread in a three-layer network. *Chaos, Solitons, Fractals* 186 (2024), 115264. https://doi.org/10.1016/j.chaos.2024.115264

[37] Trey Hedden and Jun Zhang. 2002. What do you think I think you think?: Strategic reasoning in matrix games. *Cognition* 85, 1 (2002), 1–36. https://doi.org/10.1016/S0010-0277(02)00054-9

[38] Samantha L. Heiman, Scott Claessens, Jessica D. Ayers, Diego Guevara Beltrán, Andrew Van Horn, Edward R. Hirt, Athena Aktipis, and Peter M. Todd. 2023. Descriptive norms caused increases in mask wearing during the COVID-19 pandemic. *Scientific Reports* 13, 1 (22 Jul 2023), 11856. https://doi.org/10.1038/s41598-023-38593-w

[39] Teck H. Ho, Colin F. Camerer, and Juin-Kuan Chong. 2007. Self-tuning experience weighted attraction learning in games. *Journal of Economic Theory* 133, 1 (2007), 177–198. https://doi.org/10.1016/j.jet.2005.12.008

[40] Petter Holme and M. E. J. Newman. 2006. Nonequilibrium phase transition in the coevolution of networks and opinions. *Phys. Rev. E* 74 (Nov 2006), 056108. Issue 5. https://doi.org/10.1103/PhysRevE.74.056108

[41] W. Jager, M. A. Janssen, H. J. M. De Vries, J. De Greef, and C. A. J. Vlek. 2000. Behaviour in commons dilemmas: Homo economicus and Homo psychologicus in an ecological-economic model. *Ecological Economics* 35, 3 (2000), 357–379. https://doi.org/10.1016/S0921-8009(00)00220-2

[42] Yoshihisa Kashima, Simon M. Laham, Jennifer Dix, Bianca Levis, Darlene Wong, and Melissa Wheeler. 2015. Social transmission of cultural practices and implicit attitudes. *Organizational Behavior and Human Decision Processes* 129, C (2015), 113–125. https://doi.org/10.1016/j.obhdp.2014.05.0

[43] István Z Kiss, Joel C Miller, Péter L Simon, et al. 2017. Mathematics of epidemics on networks. *Cham: Springer* 598 (2017). https://doi.org/10.1007/978-3-319-50806-1

[44] Peter Klimek and Stefan Thurner. 2013. Triadic closure dynamics drives scaling laws in social multiplex networks. *New Journal of Physics* 15, 6 (jun 2013), 063008. https://doi.org/10.1088/1367-2630/15/6/063008

[45] Joachim I. Krueger. 2007. From social projection to social behaviour. *European Review of Social Psychology* 18, 1 (2007), 1–35. https://doi.org/10.1080/10463280701284645 arXiv:https://doi.org/10.1080/10463280701284645

[46] Jinyi Kuang, Maryann G. Delea, Erik Thulin, and Cristina Bicchieri. 2020. Do descriptive norms messaging interventions backfire? Protocol for a systematic review of the boomerang effect. *Systematic Reviews* 9, 1 (24 Nov 2020), 267. https://doi.org/10.1186/s13643-020-01533-0

[47] Xueheng Li, Lucas Molleman, and Dennie van Dolder. 2021. Do Descriptive Social Norms Drive Peer Punishment? Conditional Punishment Strategies and Their Impact on Cooperation. *Evolution and Human Behavior* 42, 5 (Sep 2021), 469–479. https://doi.org/10.2139/ssrn.3571220

[48] Raúl López-Pérez. 2008. Aversion to norm-breaking: A model. *Games and Economic Behavior* 64, 1 (2008), 237–267. https://doi.org/10.1016/j.geb.2007.10.009

[49] André C. R. Martins. 2008. Continuous opinions and Discrete actions in opinion dynamics problems. *International Journal of Modern Physics C* 19, 04 (2008), 617–624. https://doi.org/10.1142/S0129183108012339 arXiv:https://doi.org/10.1142/S0129183108012339

[50] Biljana Meiske, Amalia Álvarez Benjumea, Giulia Andrighetto, and Eugenia Polizzi. 2024. Nudging punishment against sharing of fake news. *European Economic Review* 168 (2024), 104795. https://doi.org/10.1016/j.euroecorev.2024.104795

[51] Soya Miyoshi, Marko Jusup, and Petter Holme. 2021. Flexible imitation suppresses epidemics through better vaccination. *Journal of Computational Social Science* 4, 2 (01 Nov 2021), 709–720. https://doi.org/10.1007/s42001-021-00105-z

[52] Alex Moehring, Avinash Collis, Kiran Garimella, M. Amin Rahimian, Sinan Aral, and Dean Eckles. 2023. Providing normative information increases intentions to accept a COVID-19 vaccine. *Nature Communications* 14, 1 (09 Jan 2023), 126. https://doi.org/10.1038/s41467-022-35052-4

[53] K. Nyborg, J. M. Anderies, A. Dannenberg, T. Lindahl, C. Schill, M. ter, W. N. Adger, K. J. Arrow, S. Barrett, S. Carpenter, F. S. Chapin, A. S. pin, G. Daily, P. Ehrlich, C. Folke, W. Jager, N. Kautsky, S. A. Levin, O. J. Madsen, S. Polasky, M. Scheffer, B. Walker, E. U. Weber, J. Wilen, A. Xepapadeas, and A. de Zeeuw. 2016. Social norms as solutions. *Science* 354, 6308 (Oct 2016), 42–43.

[54] B. Nyhan, J. Reifler, and S. Richey. 2012. The role of social networks in influenza vaccine attitudes and intentions among college students in the Southeastern United States. *J Adolesc Health* 51, 3 (Sep 2012), 302–304. https://doi.org/10.1016/j.jadohealth.2012.02.014

[55] T. Oraby, V. Thampi, and C. T. Bauch. 2014. The influence of social norms on the dynamics of vaccinating behaviour for paediatric infectious diseases. *Proc Biol Sci* 281, 1780 (Apr 2014), 20133172. https://doi.org/10.1098/rspb.2016.1859

[56] Elinor Ostrom. 2000. Collective Action and the Evolution of Social Norms. *Journal of Economic Perspectives* 14, 3 (September 2000), 137–158. https://doi.org/10.1257/jep.14.3.137

[57] Nicola Perra, Duygu Balcan, Bruno Gonçalves, and Alessandro Vespignani. 2011. Towards a Characterization of Behavior-Disease Models. *PLOS ONE* 6, 8 (08 2011), 1–15. https://doi.org/10.1371/journal.pone.0023084

[58] Deborah Prentice and Elizabeth Levy Paluck. 2020. Engineering social change using social norms: lessons from the study of collective action. *Current Opinion in Psychology* 35 (2020), 138–142. https://doi.org/10.1016/j.copsyc.2020.06.012 Social Change (Rallies, Riots and Revolutions).

[59] S. C. Quinn, K. M. Hilyard, A. M. Jamison, J. An, G. R. Hancock, D. Musa, and V. S. Freimuth. 2017. The influence of social norms on flu vaccination among African American and White adults. *Health Educ Res* 32, 6 (Dec 2017), 473–486. https://doi.org/10.1093/her/cyx070

[60] Matthew Rabin. 1994. Cognitive dissonance and social change. *Journal of Economic Behavior and Organization* 23, 2 (1994), 177–194. https://doi.org/10.1016/0167-2681(94)90066-3

[61] David G Rand and Erez Yoeli. 2024. Descriptive norms can "backfire" in hyper-polarized contexts. *PNAS Nexus* 3, 10 (10 2024), pgae303. https://doi.org/10.1093/pnasnexus/pgae303 arXiv:https://academic.oup.com/pnasnexus/article-pdf/3/10/pgae303/59961412/pgae303.pdf

[62] J. Realpe Gomez, G. Andrighetto, L. G. Nardin, and J. A. Montoya. 2018. Balancing selfishness and norm conformity can explain human behavior in large-scale prisoner's dilemma games and can poise human groups near criticality. *Phys Rev E* 97, 4-1 (Apr 2018), 042321. https://doi.org/10.1103/PhysRevE.97.042321

[63] John Realpe-Gómez, Daniele Vilone, Giulia Andrighetto, Luis G. Nardin, and Javier A. Montoya. 2018. Learning Dynamics and Norm Psychology Supports Human Cooperation in a Large-Scale Prisoner's Dilemma on Networks. *Games* 9, 4 (2018). https://doi.org/10.3390/g9040090

[64] Timothy C. Reluga. 2010. Game Theory of Social Distancing in Response to an Epidemic. *PLOS Computational Biology* 6, 5 (05 2010), 1–9. https://doi.org/10.1371/journal.pcbi.1000793

[65] Isabel Richter, John Thøgersen, and Christian A. Klöckner. 2018. A Social Norms Intervention Going Wrong: Boomerang Effects from Descriptive Norms Information. *Sustainability* 10, 8 (2018). https://doi.org/10.3390/su10082848

[66] Omid Roozmand, Guillaume Deffuant, Giulia Andrighetto, and Mario Paolucci. 2022. Simulating collective risk management from experimental data. In *Social Simulation Conference 2022*. https://hal.science/hal-04211468

[67] Marcel Salathé, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W. Feldman, and James H. Jones. 2010. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences* 107, 51 (2010), 22020–22025. https://doi.org/10.1073/pnas.1009094108 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1009094108

[68] Angel Sánchez. 2018. Physics of human cooperation: experimental evidence and theoretical models. *Journal of Statistical Mechanics: Theory and Experiment* 2018, 2 (2018), 024001. https://doi.org/10.1088/1742-5468/aaa388

[69] P. Sheeran, A. Maki, E. Montanaro, A. Avishai-Yitshak, A. Bryan, W. M. Klein, E. Miles, and A. J. Rothman. 2016. The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: A meta-analysis. *Health Psychol* 35, 11 (Nov 2016), 1178–1188. https://doi.org/10.1037/hea0000387

[70] Benyun Shi, Guangliang Liu, Hongjun Qiu, Zhen Wang, Yizhi Ren, and Dan Chen. 2019. Exploring voluntary vaccination with bounded rationality through reinforcement learning. *Physica A: Statistical Mechanics and its Applications* 515, C (2019), 171–182. https://doi.org/10.1016/j.physa.2018.09.1

[71] S. Sinclair and J. m. 2023. Do Social Norms Influence Young People's Willingness to Take the COVID-19 Vaccine? *Health Commun* 38, 1 (Jan 2023), 152–159. https://doi.org/10.1080/10410236.2021.1937832

[72] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.

[73] Aron Szekely, Francesca Lipari, Alberto Antonioni, Mario Paolucci, Angel Sánchez, Luca Tummolini, and Giulia Andrighetto. 2021. Evidence from a long-term experiment that collective risks change social norms and promote cooperation. *Nature Communications* 12, 1 (15 Sep 2021), 5452. https://doi.org/10.1038/s41467-021-25734-w

[74] Firouzeh Taghikhah, Tatiana Filatova, and Alexey Voinov. 2021. Where Does Theory Have It Right? A Comparison of Theory-Driven and Empirical Agent Based Models. *Journal of Artificial Societies and Social Simulation* 24, 2 (2021), 4. https://doi.org/10.18564/jasss.4573

[75] L. Troyer and C. W. Younts. 1997. Whose expectations matter? The relative power of first- and second-order expectations in determining social influence. *Amer. J. Sociology* 103, 3 (1997), 692–732. https://doi.org/10.1086/231253

[76] Denis Tverskoi, Sudarsanam Babu, and Sergey Gavrilets. 2022. The spread of technological innovations: effects of psychology, culture and policy interventions. *Royal Society Open Science* 9, 6 (2022), 211833. https://doi.org/10.1098/rsos.211833 arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rsos.211833

[77] Marina Voinson, Sylvain Billiard, and Alexandra Alvergne. 2015. Beyond Rational Decision-Making: Modelling the Influence of Cognitive Biases on the Dynamics of Vaccination Coverage. *PLOS ONE* 10, 11 (11 2015), 1–17. https://doi.org/10.1371/journal.pone.0142990

[78] Eva Vriens, Giulia Andrighetto, and Luca Tummolini. 2024. Risk, sanctions and norm change: the formation and decay of social distancing norms. *Philosophical Transactions of the Royal Society B: Biological Sciences* 379, 1897 (2024), 20230035. https://doi.org/10.1098/rstb.2023.0035 arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.2023.0035

[79] Eva Vriens, Luca Tummolini, and Giulia Andrighetto. 2023. Vaccine-hesitant people misperceive the social norm of vaccination. *PNAS Nexus* 2, 5 (04 2023), pgad132. https://doi.org/10.1093/pnasnexus/pgad132 arXiv:https:/academic.oup.com/pnasnexus/article-pdf/2/5/pgad132/53408725/pgad132.pdf

[80] B. Walker, S. Barrett, S. Polasky, V. Galaz, C. Folke, G. m, F. Ackerman, K. Arrow, S. Carpenter, K. Chopra, G. Daily, P. Ehrlich, T. Hughes, N. Kautsky, S. Levin, K. G. ler, J. Shogren, J. Vincent, T. Xepapadeas, and A. de Zeeuw. 2009. Environment. Looming global-scale failures and missing institutions. *Science* 325, 5946 (Sep 2009), 1345–1346.

[81] Wei Wang, Quan-Hui Liu, Junhao Liang, Yanqing Hu, and Tao Zhou. 2019. Coevolution spreading in complex networks. *Physics Reports* 820 (2019), 1–51. https://doi.org/10.1016/j.physrep.2019.07.001 Coevolution spreading in complex networks.

[82] Z. Wang, M. A. Andrews, Z. X. Wu, L. Wang, and C. T. Bauch. 2015. Coupled disease-behavior dynamics on complex networks: A review. *Phys Life Rev* 15 (Dec 2015), 1–29. https://doi.org/10.1016/j.plrev.2015.07.006

[83] Zhishuang Wang, Chengyi Xia, Zengqiang Chen, and Guanrong Chen. 2021. Epidemic Propagation With Positive and Negative Preventive Information in Multiplex Networks. *IEEE Transactions on Cybernetics* 51, 3 (2021), 1454–1462. https://doi.org/10.1109/TCYB.2019.2960605

[84] WHO. [n.d.]. Ten threats to global health in 2019. ([n. d.]).

[85] Jan K. Woike, Sebastian Hafenbrädl, Patricia Kanngiesser, and Ralph Hertwig. 2022. The transmission game: Testing behavioral interventions in a pandemic-like simulation. *Science Advances* 8, 8 (2022), eabk0428. https://doi.org/10.1126/sciadv.abk0428 arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.abk0428

[86] Nadav Yakobovitch and Amir Grinstein. 2016. Materialism and the Boomerang Effect of Descriptive Norm Demarketing: Extension and Remedy in an Environmental Context. *Journal of Public Policy Marketing* 35, 1 (2016), 91–107. http://www.jstor.org/stable/44164870

[87] Qian Yin, Zhishuang Wang, Chengyi Xia, and Chris T. Bauch. 2022. Impact of co-evolution of negative vaccine-related information, vaccination behavior and epidemic spreading in multilayer networks. *Communications in Nonlinear Science and Numerical Simulation* 109 (2022), 106312. https://doi.org/10.1016/j.cnsns.2022.106312

[88] Haifeng Zhang and Yevgeniy Vorobeychik. 2019. Empirically grounded agent-based models of innovation diffusion: a critical review. *Artificial Intelligence Review* 52, 1 (01 Jun 2019), 707–741. https://doi.org/10.1007/s10462-017-9577-z