

**Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig**

The dynamics of opinion expression

by

Felix Gaisbauer, Eckehard Olbrich, and Sven Banisch

Preprint no.: 24

2020



The dynamics of opinion expression

Felix Gaisbauer,* Eckehard Olbrich, and Sven Banisch
Max Planck Institute for Mathematics in the Sciences
Inselstrasse 22, 04103 Leipzig

Modelling efforts in opinion dynamics have to a large extent ignored that opinion exchange between individuals can also have an effect on how willing they are to express their opinion publicly. Here, we introduce a model of public opinion expression. Two groups of agents with different opinion on an issue interact with each other, changing the willingness to express their opinion according to whether they perceive themselves as part of the majority or minority opinion. We formulate the model as a multi-group majority game and investigate the Nash equilibria. We also provide a dynamical systems perspective: Using the reinforcement learning algorithm of Q -learning, we reduce the N -agent system in a mean-field approach to two dimensions which represent the two opinion groups. This two-dimensional system is analyzed in a comprehensive bifurcation analysis of its parameters. The model identifies social-structural conditions for public opinion predominance of different groups. Among other findings, we show under which circumstances a minority can dominate public discourse.

“The actual strength of [...] different camps of opinion does not necessarily determine which view will predominate in public. An opinion can dominate in public and give rise to the pressure of isolation even if the majority of the population holds the opposing view that has come under pressure—yet does not publicly admit to holding this position.”[1]

I. INTRODUCTION

Fundamental to models of opinion dynamics is the assumption that people’s opinions are, in some way or another, influenced by the opinion of their peers. There is an extensive amount of models of opinion change in social systems (see [2–4] for reviews). While it is a plausible assumption that people who express their opinion about an issue are sensitive to approval and disapproval, feedback on the opinion need not necessarily lead to its reconsideration. It might also affect one’s willingness of opinion *expression*: The more positive (negative) the feedback, the more (less) motivated one feels to publicly express one’s opinion.

In comparison, this approach to public discourse has remained, from a modelling perspective, rather unexplored. However, it is worth to be considered: In general, people are not always willing to reveal their opinion on certain issues to others [5]. A recent study shows that only a minority of users who consume news online is also involved in sharing and discussing them [6]. Thorough research on opinion dynamics must take into account that some individuals might choose to not express their opinion publicly, which has profound effect on how others perceive the opinion climate in a social system. We will hence, in this paper, focus on a model of the *expression of*, and not the change in, opinions.

An theory of public opinion expression has already been developed around fifty years ago, with Elisabeth Noelle-Neumann’s influential ‘spiral of silence’ [1, 7]. Roughly speaking, Noelle-Neumann sees the fear of isolation as an essential drive for how humans publicly behave. Especially with respect to morally charged topics, individuals constantly and mostly sub-consciously monitor the ‘opinion landscape’ around them (they possess a “quasi-statistical sense” [1, 7]) and might refrain from expressing their opinion if they believe to be part of the minority. On the other hand, a belief to hold the majority position might encourage them to express their view. Since each individual’s decision whether to express her opinion or not influences how others perceive the opinion landscape, whose evaluation might then change accordingly, a dynamical development (for which Noelle-Neumann used the metaphor of a spiral) follows in which the seemingly dominant opinion fraction becomes more and more vocal and the perceived minority fraction becomes more and more silent. Noelle-Neumann’s spiral of silence is particularly interesting for mathematical modelling since it links a micro mechanism with a dynamical development at the macro level.

While there have been attempts to model opinion expression and specifically the spiral of silence, they are either in large parts simulative [8–12] or directed towards the effect of specific circumstances on the spiral of silence (mass media [13], social bots [9], or the long-time effect of charismatic agents [12]). Granovetter and Soong [14], and subsequently Krassa [15], employ a threshold model of opinion expression which only applies to cases in which a certain opinion is already suppressed. We aim here towards a more general, structural understanding of the dynamics of opinion expression.

We develop a model which employs an account of social influence termed social feedback theory [16]. The behavioral adjustment of agents depends solely on the social feedback they receive when they express their opinion. This affective experience-based interaction mechanism has already been shown to lead to opinion polariza-

* felix.gaisbauer@mis.mpg.de

tion in connected networks of sufficiently high modularity [17]. In the present approach, the effect of social interaction is directed towards the *willingness* of or incentive for individuals to publicly express their opinion. We investigate the structural conditions that promote or hinder opinion expression of different opinion groups. This is firstly done from a game-theoretic angle. To address questions of bounded rationality and equilibrium selection, we also develop a dynamical systems perspective, using reinforcement learning in the form of Q -learning [18]. This allows us to perform a mean-field approximation for the expected reward of the two opinion groups, which reduces the system to two dimensions.

In the following, we will first describe the baseline social structure and the two central structural parameters of the model. In section III, we represent the model as a multi-group majority game on the agent network, and investigate its Nash equilibria with respect to the structural parameters. Section IV introduces Q -learning and a subsequent two-dimensional approximation of the dynamical system. In section V we perform a bifurcation analysis for the different parameters involved. We conclude with a discussion of the results and an outlook in section VI.

II. SOCIAL-STRUCTURAL SETTING

For simplicity, we assume that there are two groups of individuals holding two different opinions on an issue. The opinion of an agent i , o_i , is referred to by either 1 or 2, depending on the group she belongs to. G_1 is the group of agents holding opinion 1, G_2 the one holding opinion 2. According to their opinion, the connections between agents are described by weighted blocks (the entries q_{11} , q_{12} , q_{21} and q_{22} in the different blocks represent the weight of every connection within that block) according to the adjacency matrix A ,

$$A = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}. \quad (1)$$

Opinion group G_1 has size N_1 and opinion group G_2 has size N_2 . The weight of an edge between any two agents of community 1 is q_{11} , and analogously q_{22} for the second community. Cross-edge weights are given by q_{12} and q_{21} . All weights $q_{11}, q_{22}, q_{12}, q_{21} \in [0, 1]$. We assume an undirected network, hence

$$q_{12} = q_{21}.$$

The weights of the edges can be interpreted as the intensity or strength of the connections: The smaller the

weight, the less strong agents notice the presence of each other. We can express the fraction of others an agent perceives to hold the same opinion by

$$f_{11} = \frac{(N_1 - 1)q_{11}}{(N_1 - 1)q_{11} + N_2q_{12}} \quad (2)$$

for agents belonging to opinion group G_1 and

$$f_{22} = \frac{(N_2 - 1)q_{22}}{(N_2 - 1)q_{22} + N_1q_{12}}, \quad (3)$$

for agents that are part of opinion group G_2 . The perceived fractions of others belonging to the other opinion group are consequently

$$f_{12} = \frac{N_2q_{12}}{(N_1 - 1)q_{11} + N_2q_{12}} \quad (4)$$

and

$$f_{21} = \frac{N_1q_{12}}{(N_2 - 1)q_{22} + N_1q_{12}}. \quad (5)$$

We now introduce the two central structural parameters, γ and δ . They are the ratios of the weighted in-group to the out-group connections for each opinion group and given by

$$\gamma = \frac{N_1 - 1}{N_2} \cdot \frac{q_{11}}{q_{12}} \quad (6)$$

and

$$\delta = \underbrace{\frac{N_2 - 1}{N_1}}_{\text{group sizes}} \cdot \underbrace{\frac{q_{22}}{q_{12}}}_{\text{weights}}, \quad (7)$$

$\gamma > 1$ or $\delta > 1$ means that the agents of one opinion group are more strongly connected (under consideration of both the group sizes and the weights) to agents of the same than of the other opinion, while $\gamma < 1$ or $\delta < 1$ indicates that agents of the opinion group are more strongly connected to agents holding a different opinion. In the following, if we say that an opinion group is internally well-connected, we mean that the structural parameter of the group is bigger than 1. With γ and δ , the above fractions (2), (3), (4) and (5) can be simplified to

$$f_{11} = \frac{\gamma}{\gamma + 1}, \quad f_{12} = \frac{1}{\gamma + 1}, \quad (8)$$

$$f_{22} = \frac{\delta}{\delta + 1}, \quad f_{21} = \frac{1}{\delta + 1}. \quad (9)$$

Alternatively, we can interpret the fractions f_{11} , f_{22} , f_{12} and f_{21} as the probabilities of interaction between agents of the different groups.¹ This interpretation will be made use of in section IV.

¹ In this case, A is interpreted as a stochastic block matrix and the weights q_{11} , q_{22} and q_{12} as probabilities of there being an edge between any two agents, depending on their group affiliation.

III. A SILENCE GAME

We now use the social structure described in section II as the setting of a ‘silence game’. The opinions of the agents are fixed according to their group affiliation and do not change. Each agent can choose one of two actions: Public expression of opinion, or silence. Their preference over the actions depends on the perception of their opinion environment. If it appears to them that they are part of a minority opinion, they become silent. If they think that they are part of the majority, they express their opinion.² But only the expressive agents shape the subjective impression of the opinion landscape of each individual. Silent ones do not contribute (they are not counted in the inequalities below). After all, silence means that the individual’s opinion is not public.

Moreover, we introduce as an additional model assumption that opinion expression does not come for free. It is costly to express one’s opinion,³ which is accounted for by a constant cost c . This constant might make more than a simple (perceived) opinion majority necessary for an agent to also have an incentive to express her opinion.

The ordinal preferences of an individual i over the actions e (for opinion expression) and s (for silence) are given as follows: The payoff or utility for opinion expression, $U_i(e, a_{-i})$ (given the actions of all others a_{-i}), is bigger than the one for silence,

$$U_i(e, a_{-i}) > U_i(s, a_{-i}),$$

if more agents in the perceived social environment of i speak out who share i ’s opinion. We can make this condition explicit by the inequality

$$\frac{\sum_{j \in G_1, j \neq i} q_{11} a_j}{(N_1 - 1)q_{11} + N_2 q_{12}} > \frac{\sum_{j \in G_2} q_{12} a_j}{(N_1 - 1)q_{11} + N_2 q_{12}} + c \quad (10)$$

if i is part of opinion group G_1 and

$$\frac{\sum_{j \in G_2, j \neq i} q_{22} a_j}{(N_2 - 1)q_{22} + N_1 q_{12}} > \frac{\sum_{j \in G_1} q_{12} a_j}{(N_2 - 1)q_{22} + N_1 q_{12}} + c \quad (11)$$

for i being part of G_2 . Both the right- and the left-hand side are normalized with respect to the overall weighted connections of agent i . Here, the actions a_j are given by $a_j = 1$ for expression and $a_j = 0$ for silence. If the two sides (10) or (11) are equal, the individual is indifferent in her preference over the actions.

A strategy profile is called a Nash equilibrium (NE) if no individual i can increase her expected reward by unilaterally deviating from the equilibrium. In our system,

the equilibrium condition is met if there is a strategy profile for which each individual that expresses herself has (10) or (11) (depending on the opinion group of the agent) satisfied, and if for each individual that is silent, the corresponding inequality is not fulfilled.

It is already visible in (10) and (11) that apart from the fact that an individual does account for her own expressed opinion in the inequality ($i \neq j$ in the sum on the left-hand side), the rest of the contributions in the inequalities are the same for all agents of one opinion group. It is also visible that if (10) or (11) is satisfied for an agent i that expresses herself, it must be satisfied for all silent individuals of her group as well: They ‘see’ one more agent expressing their opinion than i , since i does not account for herself in her evaluation of her environment. Hence, we have an additional positive term on their left-hand side. On the other hand, if the inequality is not fulfilled for a silent agent of one group, it can also not be fulfilled for an expressive one. Therefore, in a pure-strategy equilibrium, all agents of one opinion group must choose the same action.

This simplifies the inequalities above. If all agents of an opinion group act the same, (10) and (11) can be expressed in terms of the structural parameters γ and δ . Four pure-strategy NEs might be possible, depending on γ and δ . Both groups can be silent, or only one of them, but not the other, or none:

- If both groups express their opinion (we call this state (e, e) ; the first entry stands for the collective action of G_1 , the second for the action of G_2), the following conditions must be satisfied to make this state a NE:

$$\frac{(N_1 - 1)q_{11} - N_2 q_{12}}{(N_1 - 1)q_{11} + N_2 q_{12}} - c = \frac{\gamma - 1}{\gamma + 1} - c > 0, \quad (12)$$

$$\frac{(N_2 - 1)q_{22} - N_1 q_{12}}{(N_2 - 1)q_{22} + N_1 q_{12}} - c = \frac{\delta - 1}{\delta + 1} - c > 0. \quad (13)$$

- (e, s) is a NE if

$$\frac{(N_1 - 1)q_{11}}{(N_1 - 1)q_{11} + N_2 q_{12}} - c = \frac{\gamma}{\gamma + 1} - c > 0, \quad (14)$$

$$-\frac{N_1 q_{12}}{(N_2 - 1)q_{22} + N_1 q_{12}} - c = -\frac{1}{\delta + 1} - c < 0. \quad (15)$$

- (s, e) is a NE if

$$-\frac{N_2 q_{12}}{(N_1 - 1)q_{11} + N_2 q_{12}} - c = -\frac{1}{\gamma + 1} - c < 0, \quad (16)$$

$$\frac{(N_2 - 1)q_{22}}{(N_2 - 1)q_{22} + N_1 q_{12}} - c = \frac{\delta}{\delta + 1} - c > 0. \quad (17)$$

² Games with fixed, different group affiliations of agents are considered e.g. in [19] or [20].

³ We may think of the effort of typing a reply to someone in social media, or the effort of joining a demonstration for or against some issue.

⁴ We use equation (6) in the equivalence.

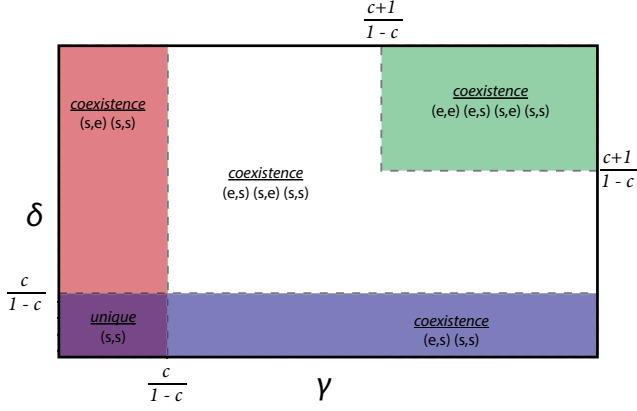


FIG. 1. The available pure-strategy Nash equilibria in different regimes of γ and δ . The equilibria are abbreviated by either e for expression or s for silence for each opinion group (the first entry is for the collective action of G_1 , the second for the one of G_2). For costs $c > 0$, γ and δ below $\frac{c}{1-c}$ will lead to a situation in which the only available Nash equilibrium is one in which no one expresses her opinion publicly. An increase in the structural parameters above this threshold leads to additional Nash equilibria in which at least one of the two opinion groups speaks out. If both γ and δ are bigger than $\frac{c+1}{1-c}$, an additional Nash equilibrium arises in which all agents express their opinion.

- (s, s) is a NE if

$$-c < 0. \quad (18)$$

The different existence regimes of the pure-strategy NEs are given in Figure 1. If γ and δ are both smaller than $\frac{c}{1-c}$, then even if all group members express their opinion and the other opinion group is silent, it is too costly (compared to the amount of connections to agents of the own opinion group) to express one's opinion and the only NE is the one in which all individuals are silent. If γ or δ or both are bigger than $\frac{c}{1-c}$, but smaller than $\frac{c+1}{1-c}$, either both opinion groups are silent or one of the groups expresses themselves, but not both: The strength of internal connections of each group are not sufficient to account for the negative influence of the other, expressive group. Not both (12) and (13) can be satisfied. Hence, this structural regime only allows public opinion predominance of one group (or complete silence).⁵ If γ and δ are both bigger than $\frac{c+1}{1-c}$, it is possible that both opinion groups express their opinion publicly at the same

time. Then, the positive influence of the in-group members still dominates, even if all out-group members are expressive as well. Hence, also (12) and (13) are satisfied.

Obviously, there are also mixed-strategy NEs. Suppose the situation is as follows: The agents of each group mix their actions uniformly such that each agent is exactly indifferent between expressing herself or staying silent. Then, no one has an incentive for action change, and we therefore have a NE. This equilibrium is, nevertheless, only metastable in the sense that it only takes one agent to increase (or decrease) her expression probability in order to make it favourable for all other agents of one opinion group to express themselves (or become silent).

γ and δ do not only depend on the number of agents holding one or another opinion. They are also influenced by the internal connection weights of agents of one opinion group. Hence, a well-connected minority group can dominate public discourse if the corresponding structural parameter is above the threshold of $\frac{c}{1-c}$. But while the regimes of different NEs in Figure 1 are displayed correctly, it might give the impression that γ or δ are parameters that can be tuned by simply increasing the weight of a connection between two agents of the same group, that is, q_{11} or q_{22} (all other parameters fixed). That is not the case. Some numerical minorities cannot be balanced by increasing internal connections since q_{11} and q_{22} are bounded by 1. If there are too few agents in one opinion group, even setting q_{11} or q_{22} to 1 will not be able to elevate γ or δ above a certain threshold. This is made visible in Figure 2. The figure shows the different existence regimes of the NEs for different combinations of internal connection weights q_{11} and q_{22} and partitions of a total of $N = 100$ agents between groups G_1 and G_2 . q_{12} and c are fixed. Each point in the plot stands for a combination of the number of agents in opinion group G_1 , N_1 , and the in-group connection weights q_{11} , out of which one can compute the value of γ . The lines of constant γ are plotted in red. Since the overall number of agents $N = 100$ is fixed, N_2 is not independent and determined by the choice of N_1 by $N - N_1$. If we just assume that $q_{22} = q_{11}$, each point in the plot at the same time represents also a combination of the relevant parameters of opinion group G_2 out of which one can compute δ . Curves of constant δ are the blue lines and symmetrical to the γ -curves with respect to $N_1 = 50$.

A vertical line in the plot, e.g. at $N_1 = 20$, can be interpreted as follows: Each constant γ or δ value that it intersects on its way to $q_{11} = q_{22} = 1$ is reachable for this partition of agents in the two groups if q_{11} and q_{22} are tuned accordingly. But if there is no intersection for a specific γ or δ , then even if the internal connection weights are maximized, the structural strength of the respective group cannot reach that value due to their limited group size. For $N_1 = 20$, a state in which both opinion groups are expressing themselves (the upper right, green area in Figure 1) cannot be reached since opinion group G_1 has too few agents to produce a γ high enough to satisfy (12). In general, there are numerical thresh-

⁵ If either only the conditions for (e, s) or only for (s, e) are satisfied, it is clear which opinion will dominate publicly (if any). If both are satisfied, the situation becomes more interesting in the sense that it depends on the initial conditions and the dynamical development of the system which opinion will predominate. We will approach these issues in sections IV and V.

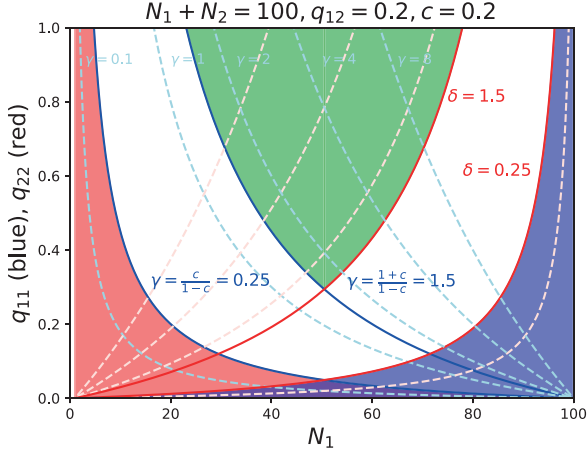


FIG. 2. The constant γ - and δ -curves for $N = 100$ agents. They are plotted with respect to N_1 , $N_2 = N - N_1$, and $q_{11} = q_{22}$. Each blue curve stands for a combination of the number of opinion group members N_1 and internal connection weights q_{11} that yields a constant value of the structural parameters γ , each red one for a combination of N_2 and q_{22} that produces constant δ . The color-coding for the different Nash equilibrium regions is analogous to Figure 1. It is visible that the numerical minority of an opinion group cannot always be compensated by increasing q_{11} (or q_{22}), the weight of a connection between two agents of the same opinion group. Moreover, the fixed γ - and δ -curves are symmetric with respect to $N_1 = N_2 = 50$, where they intersect. (For better readability, the dashed δ -curves have not been labelled. They correspond to their γ -counterparts.)

olds (dependent on the costs c , the cross-group connection weight q_{12} and the overall number of agents N) below which reaching a state in which both group express themselves or in which the own group becomes dominant becomes impossible from a game-theoretic perspective. The game-theoretic approach hence can give (all other parameters fixed) limits for the effect of group-internal coordination in the form of internal cohesion on public discourse.

IV. Q-LEARNING AND A DYNAMICAL SYSTEMS PERSPECTIVE

While we are able to determine the Nash equilibria of the system, the game-theoretical point of view does not answer questions of equilibrium selection or the effects of bounded rationality. In this section, we will introduce a dynamical systems perspective to approach those questions.

We posit a simple interaction mechanism between the agents on the network of section II. It is given as follows: If an agent expresses her opinion, she will be paired with a random neighbor. Now, the fractions f_{11} , f_{12} , f_{21} , and f_{22} correspond to the *probability* of meeting a neighbor of a certain opinion group given the own opinion group

of an agent. The neighbor then gives (if she also is in an expressive state) social feedback to the agent, either agreement or disagreement, which will contribute to the agent's impression of her opinion environment. Put in an algorithmic way:

1. A random agent is selected.
2. If willing to speak out, the agent expresses her opinion to a random neighbor at cost c .
3. If the neighbor is also willing to speak out, she gives feedback on the agent's opinion.
4. According to the feedback, the agent will become more/less willing to speak out.

As in [17], we will describe the development of the system as reinforcement learning dynamics, more specifically, as dynamics induced by Q -learning. In Q -learning, the reinforcement mechanism that updates the agent's willingness to express her opinion is given by

$$Q_i^{t+1} = (1 - \alpha)Q_i^t + \alpha r_i^t, \quad (19)$$

where r_i^t is the reward for agent i at time step t upon expression

$$r_i^t = \begin{cases} -c & \text{for random neighbor being silent,} \\ -1 - c & \text{for disagreeing random neighbor,} \\ 1 - c & \text{for agreeing random neighbor.} \end{cases} \quad (20)$$

The Q -function is expected to converge to the reward over time.⁶ The probability of expression is a function of the value of Q_i . We assume here a Boltzmann action selection mechanism, i.e. the probability of expression of agent i is given by

$$p_i^t = \frac{1}{1 + e^{-\beta Q_i^t}}, \quad (21)$$

the probability of staying silent by $1 - p_i^t$. If $\beta = 0$, the action choice of the agent is completely independent of the Q -values and randomized. For increasing β , the agent becomes more sensitive in her action selection towards her current evaluation of her local opinion environment. Then, a positive Q -value indicates that it is more likely for her to express herself than not, while a negative one indicates the opposite. If $\beta \rightarrow \infty$, the probabilities of the actions become deterministic.

The expected reward for agent i upon opinion expression is given by either (if i belongs to opinion group G_1)

$$\begin{aligned} \mathbb{E}_p[r_i^t] = & -c + f_{11} \frac{1}{N_1 - 1} \sum_{\substack{j \in G_1 \\ j \neq i}} \frac{1}{1 + e^{-\beta Q_j^t}} - \\ & f_{12} \frac{1}{N_2} \sum_{j \in G_2} \frac{1}{1 + e^{-\beta Q_j^t}}, \end{aligned} \quad (22)$$

⁶ (19) describes Q learning for myopic agents, i.e. with discount factor 0.

or (if i belongs to opinion group G_2)

$$\begin{aligned} \mathbb{E}_p[r_i^t] = & -c + f_{22} \frac{1}{N_2 - 1} \sum_{\substack{j \in G_2 \\ j \neq i}} \frac{1}{1 + e^{-\beta Q_j^t}} - \\ & f_{21} \frac{1}{N_1} \sum_{j \in G_1} \frac{1}{1 + e^{-\beta Q_j^t}}. \end{aligned} \quad (23)$$

We follow [18], where Q-learning in two-player two-action games is investigated, and take the continuous-time limit of the Q-learning equation (19). We divide time there into intervals δt . We replace $t + 1$ with $t + \delta t$ and α with $\alpha' \delta t$. This yields

$$Q_i(t + \delta t) - Q_i(t) = \alpha' \delta t (r_i(t) - Q_i(t))$$

and hence

$$\dot{Q}_i = \alpha' (r_i(t) - Q_i(t)). \quad (24)$$

Over time, the difference of the largest and the lowest Q -value of an opinion group decays at least exponentially in expectation (see the Appendix for the estimation):

$$\frac{d}{dt}(Q_{i \in G_1}^{\max} - Q_{i \in G_1}^{\min}) \leq -\alpha' (Q_{i \in G_1}^{\max} - Q_{i \in G_1}^{\min}),$$

$$\frac{d}{dt}(Q_{i \in G_2}^{\max} - Q_{i \in G_2}^{\min}) \leq -\alpha' (Q_{i \in G_2}^{\max} - Q_{i \in G_2}^{\min}).$$

That is, the Q -values of the agents of one group are expected to converge over time. This allows us to employ a mean-field approximation for the expected reward of the two opinion groups: We introduce the average Q -values for each opinion group⁷

$$Q_1(t) = \frac{1}{N_1} \sum_{i \in G_1} Q_i(t), \quad Q_2(t) = \frac{1}{N_2} \sum_{i \in G_2} Q_i(t). \quad (25)$$

This means that we do not distinguish any more between the agents of the respective opinion groups. We assign them the average of their group's Q -value. This simplification will have an effect on the probability of opinion expression for the individuals. Instead of averaging over each groups probability of expression, we simply insert the averaged Q -values into the equation:

$$\frac{1}{N_1} \sum_{j \in G_1} \frac{1}{1 + e^{-\beta Q_j(t)}} \rightarrow \frac{1}{1 + e^{-\beta Q_1(t)}} = p_1(t), \quad (26)$$

$$\frac{1}{N_2} \sum_{j \in G_2} \frac{1}{1 + e^{-\beta Q_j(t)}} \rightarrow \frac{1}{1 + e^{-\beta Q_2(t)}} = p_2(t). \quad (27)$$

⁷ Note the slight abuse of notation here: From now on, the index of Q and p will not indicate single individuals any more, but the average Q -value and the corresponding expression probability of the different opinion groups.

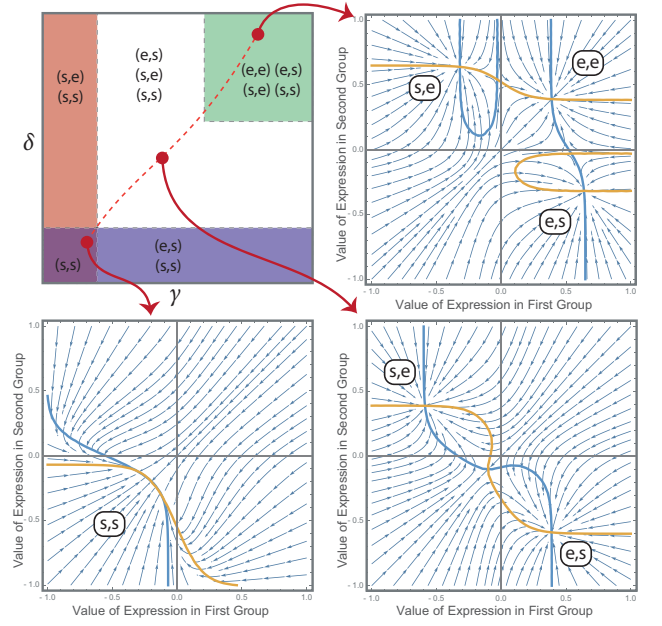


FIG. 3. Three phase portraits of the Q_1 - (x-axis) and Q_2 - values (y-axis) of the two-dimensional system for different configurations of γ and δ . We have $c = 0.1$, $\beta = 10$, and structural parameters $\gamma = \delta = 0.1$ (bottom left), $\gamma = \delta = 1$ (bottom right), and $\gamma = \delta = 3$ (top right). The yellow and blue lines in the phase portraits are the isoclines of the equations for Q_1 and Q_2 . The fixed points are located at their intersections.

The expected reward for the different opinion groups are given by the equations⁸

$$\mathbb{E}_p[r_1(t)] = -c + \frac{\gamma}{\gamma + 1} p_1(t) - \frac{1}{\gamma + 1} p_2(t), \quad (28)$$

$$\mathbb{E}_p[r_2(t)] = -c + \frac{\delta}{\delta + 1} p_2(t) - \frac{1}{\delta + 1} p_1(t), \quad (29)$$

where the probabilities of expression for each group are $p_1(t)$ and $p_2(t)$, and it is not distinguished any more between the individuals.

We can therefore write our two-dimensional formulation as follows:

$$\dot{Q}_1(t) = \alpha' (-c + \frac{\gamma}{\gamma + 1} p_1(t) - \frac{1}{\gamma + 1} p_2(t) - Q_1(t)), \quad (30)$$

$$\dot{Q}_2(t) = \alpha' (-c + \frac{\delta}{\delta + 1} p_2(t) - \frac{1}{\delta + 1} p_1(t) - Q_2(t)). \quad (31)$$

According to equations (30) and (31), we can produce a phase portrait of the system including its trajectories

⁸ f_{11} , f_{12} , f_{21} , and f_{22} have been replaced according to equations (8) and (9) with $\frac{\gamma}{\gamma+1}$, $\frac{1}{\gamma+1}$, $\frac{\delta}{\delta+1}$, and $\frac{1}{\delta+1}$.

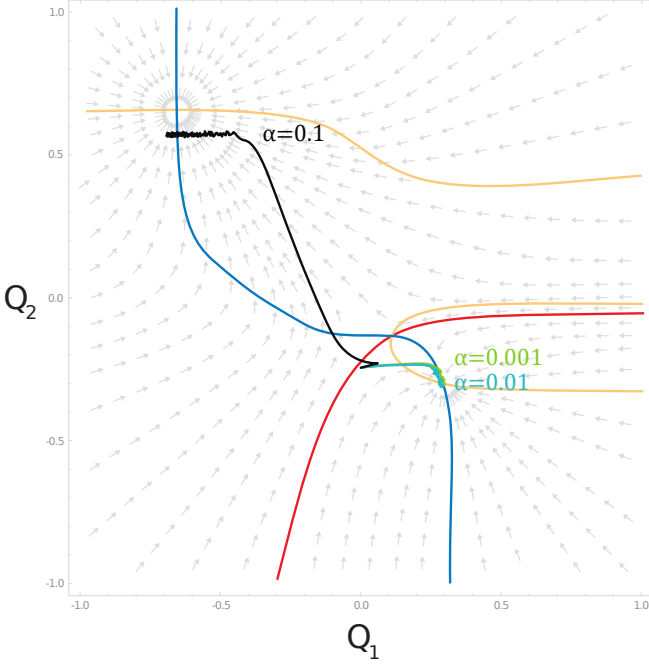


FIG. 4. The trajectories of the Q -values in simulations, averaged over 50 runs with $N \cdot 10^5$ steps, for different values of α with a starting point close to the border (red line) of the two basins of attraction of the two stable fixed points. The starting Q -values were $Q_{i \in G_1} = 0$, $Q_{i \in G_2} = -0.25$. There were $N = 200$ agents, 100 of each opinion group, and $c = 0.1$, $q_{11} = 0.04$, $q_{12} = 0.05$, and $q_{22} = 0.15$. A relatively big $\alpha = 0.1$ makes the trajectory leave the lower right basin of attraction of the two-dimensional system (black trajectory). Due to the high α , the fixed point of the other basin of attraction is also missed by some margin. The lower α , the closer the trajectories get to the fixed point and the more probable it is that they will stay in the basin predicted by the two-dimensional approximation. For $\alpha = 0.01$ (turquoise) and $\alpha = 0.001$ (light green), the trajectories run towards the predicted fixed point. The yellow and blue lines are the isoclines of the equations for Q_1 and Q_2 . The fixed points are located at their intersections.

and fixed points for given exploration rate β , structural parameters γ and δ , and costs of expression c . An example of how the phase portraits change with γ and δ is given in Figure 3.

There, it is visible that the stable fixed points of the system include basins of attraction, that is, regimes of values of Q_1 and Q_2 for which the system is expected to end up in those fixed points. The basins of attraction in the two-dimensional approximation correspond exactly to those of the stochastic N -agent system in the limit $\alpha \rightarrow 0$. For larger α , both fixed points and basins of attraction do not necessarily correspond to the two-dimensional approximation. We show averages over simulation runs for different values of α in Figure 4.

V. BIFURCATION AND STABILITY ANALYSIS

In order to find the fixed points of Q_1 and Q_2 , we set (30) and (31) to 0, solve (30) for Q_2 and insert it into (31), which yields:

$$Q_2 = -\frac{1}{\beta} \ln \left(\frac{1}{\frac{\gamma}{1+e^{-\beta Q_1}} - (\gamma+1)(Q_1+c)} - 1 \right) \quad (32)$$

$$\frac{\delta}{\delta+1} \left(\frac{\gamma}{1+e^{-\beta Q_1}} - (\gamma+1)(Q_1+c) \right) - \frac{1}{\delta+1} \frac{1}{1+e^{-\beta Q_1}} + \frac{1}{\beta} \ln \left(\frac{1}{\frac{\gamma}{1+e^{-\beta Q_1}} - (\gamma+1)(Q_1+c)} - 1 \right) - c = 0 \quad (33)$$

Equation (33) now gives us the Q_1 -value of the fixed points of the system, with which we can calculate the corresponding Q_2 -value by equation (32). In essence, the fixed points depend on four parameters: β , γ , δ , and c . We will carry out a bifurcation analysis of these parameters in the following.

After having solved equations (33) and (32) for Q_1 and Q_2 , we can assess the stability of the respective fixed points by calculating the eigenvalues of their Jacobian; two negative (real parts of the) eigenvalues indicate a stable attractor. In the following, we analyze the bifurcation structure of the system depending on the different types of parameters in the system.

A. Structural power

The parameter γ describes the ratio of internal versus external connectedness of G_1 . $\gamma > 1$ means that on average each member of G_1 is connected to more agents of the own than of the other opinion group. (Everything stated in this paragraph applies equivalently to δ , which is just the parameter for the ratio of internal versus external connectedness of the other group.)

As is visible in Figure 5, for small γ (< 0.5), given $\beta = 10$, $\delta = 2.36$ (that is, a quite well-connected opposite opinion group) and $c = 0.1$, there is only one (stable) fixed point with negative Q_1 -value and positive Q_2 . While γ grows, a saddle-node bifurcation occurs such that one stable and one unstable fixed point appear for positive Q_1 and negative Q_2 . Another saddle-node bifurcation occurs at around $\gamma = 2$; and for $\gamma > 4.2$, the low- Q_1 fixed points disappear in another saddle-node.

How can this be interpreted? In essence, an opinion community that is not well-connected internally ($\gamma < 0.5$) will be driven into silence (a Q -value much lower than 0) by the opposite opinion group that is internally more cohesive. With increasing γ , that is, increasing internal connectedness, other fixed points appear in which the former group is expressive. To be precise, the Q -values here are only indicative of probabilities of opinion expression according to the Boltzmann action selection

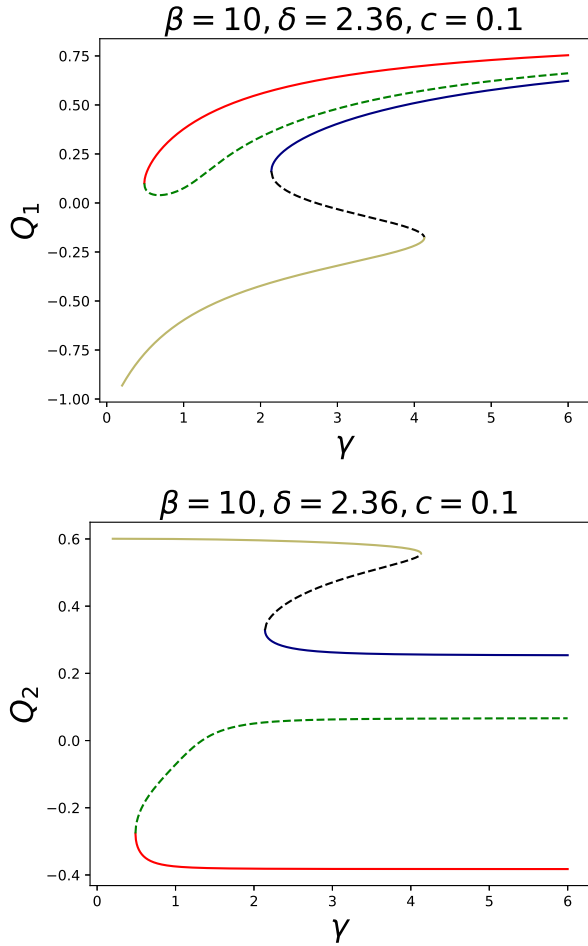


FIG. 5. The development of the Q_1 - (top) and Q_2 -value (bottom) of the fixed points with γ given β , relatively high δ , and c . The colors of the curves in the two plots indicate the different fixed point pairs of Q_1 and Q_2 . A dashed line indicates an unstable fixed point, a continuous one a stable fixed point. It is visible in the plots that a poorly connected opinion group G_1 ($\gamma < 0.5$) will be driven into silence by the other group (beige curve). With increasing in-group connectivity, fixed points arise for which G_1 expresses their opinion in two saddle-node bifurcations (red for an (e, s) -equilibrium and blue for (e, e)). For $\gamma > 4.5$, G_1 is so well-connected that the equilibrium disappears in which the group is silent.

which depends on Q . If Q is smaller than 0, the probability of expression is smaller than the probability of staying silent. In the following, if we say that one opinion group is expressive, we mean that they have a Q -value bigger than 0 which makes their probability of expression higher than that of silence. With a further increase of γ , the stable fixed point for which only the opposite opinion group is expressive disappears and we remain with three fixed points (the middle one unstable), for which either the first opinion group is ‘loud’ alone or both groups express their opinions. Hence, G_1 is now too cohesive to be driven into silence by the other group. Increased internal

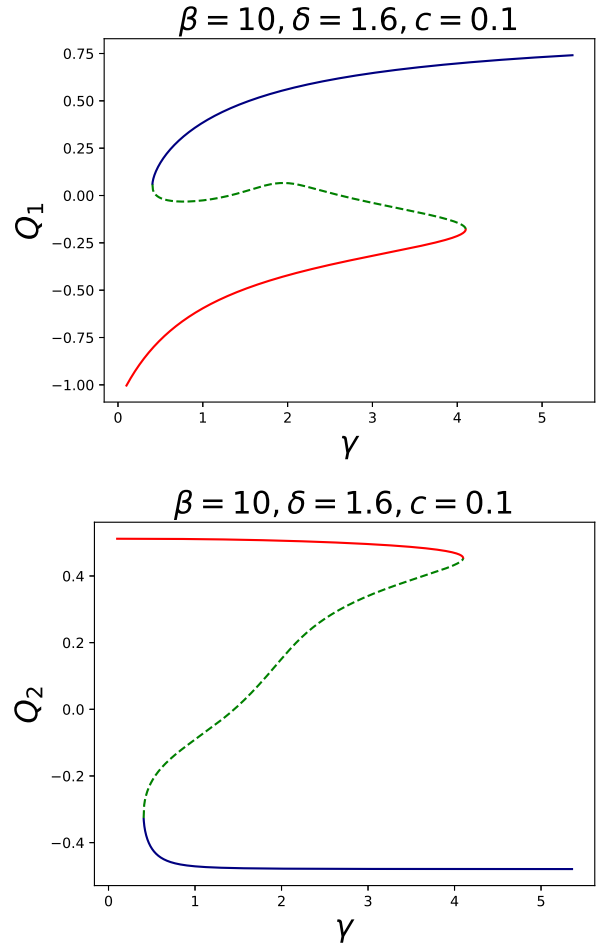


FIG. 6. The development of Q_1 - and Q_2 -fixed points with γ given β , moderate δ and c . For $\gamma < 0.4$, only group G_2 is expressive. A second fixed point arises for higher γ in which G_1 is predominating public discourse. There is no fixed point in which both groups are expressive.

cohesion of one opinion group can hence have the effect that this group, which is not necessarily a majority, will dominate public discourse.

A lower δ -value (e.g. $\delta = 1.6$) leads to a reduction in available fixed points (Figure 6) such that only two saddle-node bifurcations occur and at high γ only one fixed point remains in which the first opinion group is expressive.

B. Costs

The costs for opinion expression have a profound impact on the fixed points of the system. If opinion expression is very ‘expensive,’ (in Figure 7: $c > 0.4$), there is only one fixed point in the system for which both opinion groups stay silent. For decreasing costs, two pairs of fixed points arise in a saddle-node bifurcation. Each of

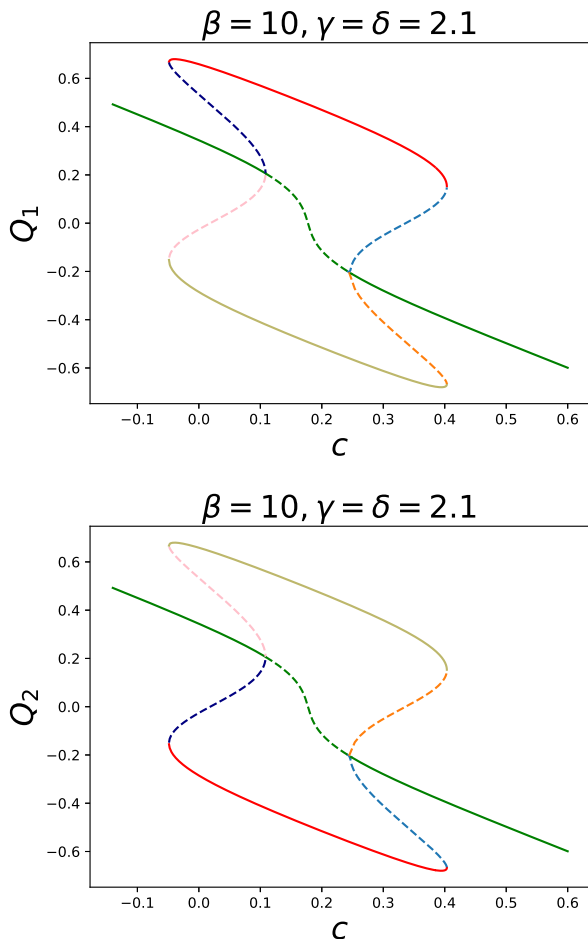


FIG. 7. The development of the fixed points with c given β and $\gamma = \delta$. The two are symmetric since c is the same for both and has the same impact on both groups if they also have identical structural parameters. If expression is costly, everyone is silent, if it has negative costs, everyone speaks out.

the pairs corresponds to a situation in which one opinion group is expressive, while the other is silent (in Figure 7, we have identical values for γ and δ). The fixed point in which both opinion groups are silent becomes unstable with decreasing c in a pitchfork bifurcation. Below $c = 0.1$, another pitchfork bifurcation arises for which the stable fixed point now corresponds to a state in which both groups are expressing their opinion. Costs can also be negative: Then, the individuals might be intrinsically motivated or externally encouraged to speak out.⁹ For sufficiently negative costs (in the case of Figure 7: $c < -0.05$), only one fixed point exists: Everyone

has an incentive to speak out, at least for internally well-connected opinion groups. The fixed points for which only one of the groups is expressive disappear in two saddle-nodes.

C. Asymmetric costs

The model allows us to also assign different costs to each opinion group, such that $c_1 \neq c_2$. Internal motivation for a cause, for example, can be an incentive to speak out and might even be indicated by negative costs (that is, an urge to express one's opinion). Moreover, there might be biases in the infrastructures on which debate takes place such that it takes more effort for one group to speak out than for the other.¹⁰

The bifurcation in Figure 8 (for the case of two internally well-connected opinion groups) illustrates the effect that different expression costs in the populations exhibit on public discourse. In Figure 8, a bifurcation over c_1 is shown. Negative costs for opinion expression in opinion group G_1 yield two stable equilibria in which opinion group 1 is expressive, either together with opinion group G_2 or alone. With increasing costs, a stable fixed point arises in a saddle node for which G_1 is silent (at $c_1 \approx 0$), while G_2 is expressive. At $c_1 \approx 0.15$ and at $c_1 \approx 0.4$, the two fixed points for which G_1 expresses opinion disappear. For costs that high, opinion group G_1 will not be publicly audible any more. Asymmetric costs can hence drive certain opinion groups into silence.

D. Exploration rate

The parameter β determines how sensitive agents are in their actions towards the current evaluation of their expected reward. A high β -value indicates a choice of the agent similar to a best response to their current evaluation of the expected reward, while $\beta = 0$ means that each available action is chosen with equal probability.

As is visible in Figure 9, for very low β , there is only one fixed point available with a very low Q -value for both opinion groups. With β (≈ 5), further fixed points arise in a supercritical pitchfork bifurcation, and then, at $\beta > 6$, another (now subcritical) pitchfork bifurcation arises, such that we arrive at three stable fixed points (one in which both groups are in an expressive mode, and one for opinion dominance for each group) and two unstable ones in-between. Hence, if the action selections is close to a best response, we get more possible equilibria in the system. In the intermediate region, we have a situation in which only one of the two groups can be expressive, despite them both being internally well-connected.

⁹ Ideas such as e.g. free speech might have such an effect: People then see it as their duty to voice their opinion, *especially* if it does not conform to the apparent majority.

¹⁰ One may think here about online platforms whose design favours engagement of certain demographic groups or states that encourage or try to prevent certain opinion groups to speak out.

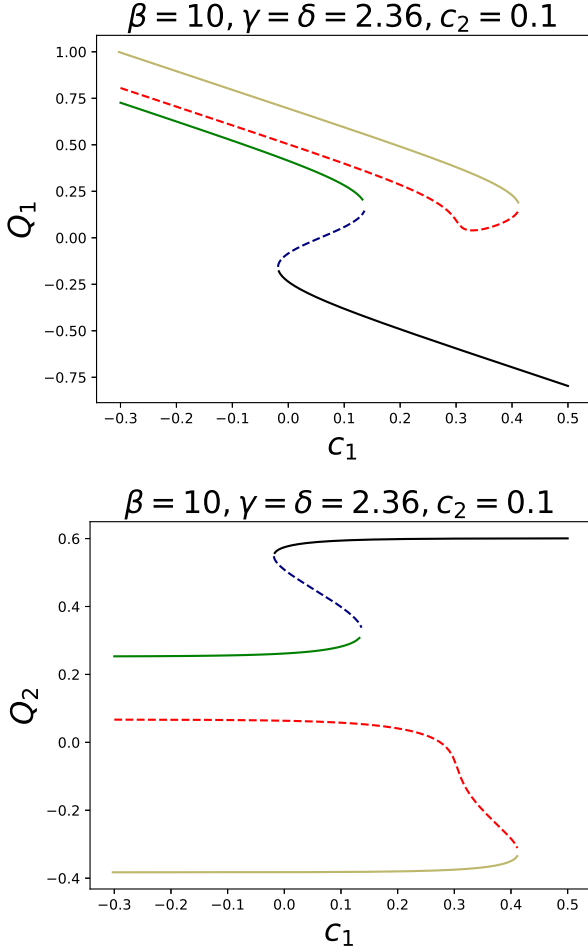


FIG. 8. Fixed-point development with c_1 independent of c_2 , given $\beta = 10$, $\gamma = \delta = 2.36$, $c_2 = 0.1$. Strongly negative c_1 corresponds to a strong motivational disposition (or many hard cores) in the opinion group to express their opinion. There, only fixed points in which this opinion group is expressive exist. For decreasing motivation (fewer hard cores), fixed points arise in which the second opinion group is the only expressive one.

VI. DISCUSSION AND OUTLOOK

a. The spiral of silence and beyond. The present model provides a structural view on collective opinion expression. It reproduces the counterintuitive result postulated by Noelle-Neumann in her theory of the spiral of silence [1, 7], namely the possibility of the public dominance of a minority opinion. While the influence of mass media has been stressed in many publications concerning the spiral of silence, we show that no mass media is needed for this effect. Being an internally well-connected community alone can be enough to gain public opinion predominance. This finding gains traction in light of the advent of social media, which facilitated communication among like-minded people and decentralized information distribution. Apart from that, the present approach also

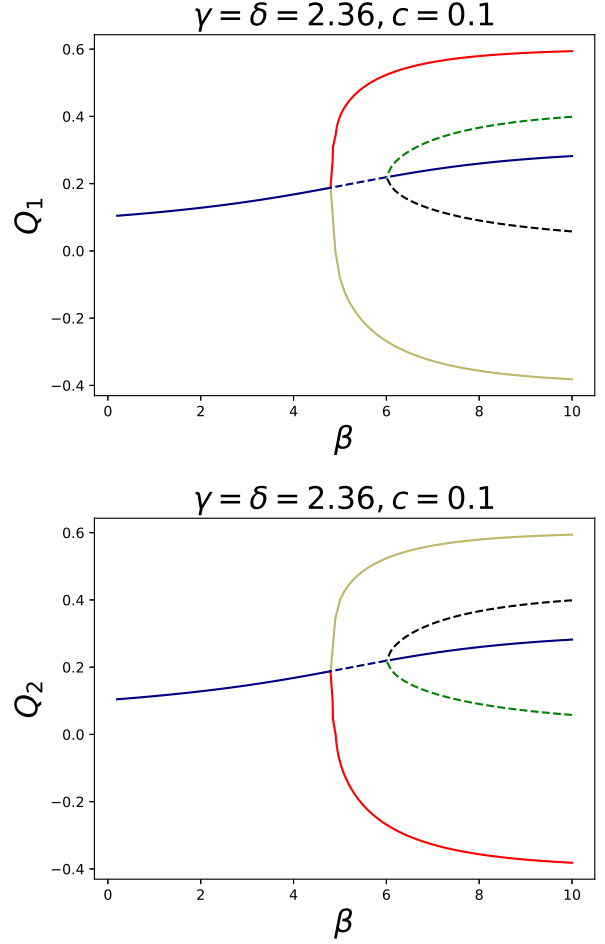


FIG. 9. The development of the fixed points with β given $c = 0.1$ and $\gamma = \delta = 2.36$. Since γ and δ are the same, the plots are symmetric.

provides conditions for the ‘overcoming’ of the spiral of silence (in the sense that both groups express their opinion publicly), for which the numerical proportions do not necessarily have to change. The increase in internal cohesion of the different opinion groups can be sufficient. On the other hand, it is also shown that if the minority is too small, even maximum internal cohesion cannot heave the minority opinion into public predominance (see again Figure 2).

b. Perception biases. In [8], the effect of the ego-network size, that is, the (average) number connections of the agents, on the occurrence of the spiral of silence was investigated. It was concluded that an increase in network density makes it more probable that one opinion group does not speak out publicly. In our work, we show that more density might even have the opposite effect. It depends on *where* the additional connections are made: If new connections are guided by homophily, such that the opinion blocks become more cohesive, the spiral of silence might even be overcome (see path (i)

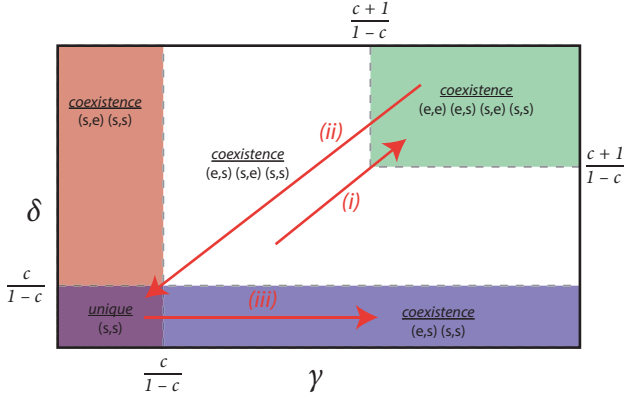


FIG. 10. Illustration of the transitions between the game-theoretic equilibrium regions for (i) stronger internal cohesion of both opinion groups ('echo chambers'), (ii) less internal cohesion of both (heterophilious connections), and (iii) stronger internal cohesion for only one opinion group ('#metoo').

in Figure 10). We then arrive at a structure similar to 'echo chambers,' in which only the voices affirming one's own view are heard and the others are blocked out (see [21] for a contribution linking opinion dynamics to the emergence of echo chambers). If the additional connections are made between the opinion blocks, both γ and δ decrease, which might make it more probable that the individuals have a more realistic picture of the overall opinion landscape. Then, the spiral effect is indeed more probable. But if the cross-group connections grow even further, both opinion groups misjudge their proportion to their own disadvantage, such that no group speaks out if there are costs associated to opinion expression (path (ii) in Figure 10). The structure of the social contacts alone is already sufficient to cause misjudgements about opinion proportions in a social system. This is closely linked to more general accounts of perception biases [22].

c. Critical mass. Furthermore, the model links to studies dealing with tipping points in social systems and the necessary numerical allocations, depending on the network structures. This has e.g. been analyzed for social conventions [23]. If the social network of individuals is structured in opinion blocks, there is a hard numerical limit for the overcoming of a state in which one opinion is dominating publicly. For example, for a cross-opinion connection probability of 0.2 (as in Figure 2), the state in which both opinion camps are expressive cannot be reached if the minority makes up less than 22% of the population.

d. Limits and outlook. While we have stressed the generality of this work, we want to emphasize its limits as well: The homogeneous network structure of opinion blocks is not particularly realistic. Real social networks are rather heterogeneous, with well-connected and very active hubs and more 'remote' individuals. Nevertheless, weighted (or stochastic [24]) blocks can serve as a baseline for mathematical accessibility.

Moreover, this work is concerned with one way of reacting on social feedback, namely, the change in willingness to express one's opinion. Change in opinion is not included. It is probable that these phenomena take place on different time scales. Also, the social environments prompting opinion change might be different from the ones in which opinion predominance is fought for. In demonstrations, if two opinion camps meet each other, the main objective might not be information exchange or the need to convince each other, but to gain public audibility. Hence, a combination of models of opinion change and opinion expression might be in order in a multi-layer network approach, in which opinion formation and the competition for public opinion predominance take place on possibly different but interdependent network structures.

While there are plenty of studies on experimental evidence for the micromechanisms grounding the spiral of silence (see [5] for a review), we are also seeking a more systematic larger-scale view on collective phenomena of opinion expression, which are closely related to the parameters γ and δ in the model. A very prominent example of emerging collective opinion expression online, for which this model provides an explanation, is the Twitter-hashtag '#metoo' and the subsequent movement against sexual harassment and sexual assault: Women found a device (in this case, a hashtag) that allowed them to find and connect to people who had experienced the same, and also to people who supported them. And all of a sudden, it was easier for them to speak out (path (iii) in Figure 10). Measurements are an intricate task here: The networks one constructs out of interactions between individuals are only the networks *of interaction*, that is, of only one part of the actions one wants to observe. Silent individuals do usually not show up in such networks since they are not involved in an observable way.

In conclusion, we develop a model of opinion expression which allows the investigation of how social structures can prevent or promote public opinion expression of different opinion groups. This approach allows direct connection to an influential theory of the social sciences, the spiral of silence [1, 7]. We approach the model both from a game-theoretic and from a dynamical systems perspective and show how the public audibility of certain opinions depends on the sensitivity of the agents towards their current evaluation of expected reward, the structural cohesion of the opinion groups and the costs for opinion expression.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 732942 (www.ODYCCEUS.eu). We are grateful for the repeated discussion of the ideas described in the paper with Wolfram Barfuss, Sharwin Rezagholi, Florentin Münch,

Armin Pournaki, and with the members of ODYCEUS.

Appendix A: Expected decrease of the difference in Q -values

We carry out the estimation for opinion group G_1 , but the analogue holds for opinion group G_2 . We can give

$$\begin{aligned} \dot{Q}_{i \in G_1}^{\max} = & \alpha' \left(\frac{\gamma}{\gamma+1} \frac{1}{N_1-1} \sum_{\substack{j \in G_1 \\ j \neq i}} \frac{1}{1+e^{-\beta Q_j}} - \frac{1}{\gamma+1} \frac{1}{N_2} \sum_{j \in G_2} \frac{1}{1+e^{-\beta Q_j}} - Q_{i \in G_1}^{\max} - c \right) \leq \\ & \alpha' \left(\frac{\gamma}{\gamma+1} \frac{1}{N_1} \left(\sum_{\substack{j \in G_1 \\ j \neq i}} \frac{1}{1+e^{-\beta Q_j}} + \frac{1}{1+e^{-\beta Q_{i \in G_1}^{\max}}} \right) - \frac{1}{\gamma+1} \frac{1}{N_2} \sum_{j \in G_2} \frac{1}{1+e^{-\beta Q_j}} - Q_{i \in G_1}^{\max} - c \right), \end{aligned} \quad (A1)$$

$$\begin{aligned} \dot{Q}_{i \in G_1}^{\min} = & \alpha' \left(\frac{\gamma}{\gamma+1} \frac{1}{N_1-1} \sum_{\substack{j \in G_1 \\ j \neq i}} \frac{1}{1+e^{-\beta Q_j}} - \frac{1}{\gamma+1} \frac{1}{N_2} \sum_{j \in G_2} \frac{1}{1+e^{-\beta Q_j}} - Q_{i \in G_1}^{\min} - c \right) \geq \\ & \alpha' \left(\frac{\gamma}{\gamma+1} \frac{1}{N_1} \left(\sum_{\substack{j \in G_1 \\ j \neq i}} \frac{1}{1+e^{-\beta Q_j}} + \frac{1}{1+e^{-\beta Q_{i \in G_1}^{\min}}} \right) - \frac{1}{\gamma+1} \frac{1}{N_2} \sum_{j \in G_2} \frac{1}{1+e^{-\beta Q_j}} - Q_{i \in G_1}^{\min} - c \right). \end{aligned} \quad (A2)$$

If we now look at the change in time in the difference of $Q_{i \in N_1}^{\max}$ and $Q_{i \in N_1}^{\min}$, we can conclude by the above inequalities that the difference decreases at least exponentially in expectation by subtracting the right hand-sides of (A1)

an upper bound for the change in Q -value for the agent with the maximum Q -value of the group, $\dot{Q}_{i \in G_1}^{\max}$, and a lower bound for the change in Q -value for the agent with the minimum Q -value of the group, $\dot{Q}_{i \in G_1}^{\min}$ due to the monotonicity of the function $\frac{1}{1+e^{-x}}$:

and (A2).

$$\frac{d}{dt} (Q_{i \in G_1}^{\max} - Q_{i \in G_1}^{\min}) \leq -\alpha' (Q_{i \in G_1}^{\max} - Q_{i \in G_1}^{\min}). \quad (A3)$$

The analogue holds for opinion group G_2 :

$$\frac{d}{dt} (Q_{i \in G_2}^{\max} - Q_{i \in G_2}^{\min}) \leq -\alpha' (Q_{i \in G_2}^{\max} - Q_{i \in G_2}^{\min}). \quad (A4)$$

-
- [1] E. Noelle-Neumann and T. Petersen, The spiral of silence and the social nature of man, in *Handbook of political communication research* (Routledge, 2004) pp. 357–374.
 - [2] J. Lorenz, Continuous opinion dynamics under bounded confidence: A survey, *International Journal of Modern Physics C* **18**, 1819 (2007).
 - [3] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics, *Reviews of Modern Physics* **81**, 591 (2009).
 - [4] A. Flache, M. Mäs, T. Feliciani, E. Chattoe-Brown, G. Deffuant, S. Huet, and J. Lorenz, Models of social influence: Towards the next frontiers, *Journal of Artificial Societies and Social Simulation* **20**, 2 (2017).
 - [5] J. Matthes, J. Knoll, and C. von Sikorski, The “spiral of silence” revisited: A meta-analysis on the relationship between perceptions of opinion support and political opinion expression, *Communication Research* **45**, 3 (2018).
 - [6] A. Kalogeropoulos, S. Negro, I. Picone, and R. K. Nielsen, Who shares and comments on news?: A cross-national comparative analysis of online and social media participation, *Social Media and Society* **3** (2017).
 - [7] E. Noelle-Neumann, The spiral of silence. A theory of public opinion, *Journal of Communication* **24**, 43 (1974).
 - [8] D. Sohn and N. Geidner, Collective Dynamics of the Spiral of Silence: The Role of Ego-Network Size, *International Journal of Public Opinion Research* **28**, 25 (2015).
 - [9] B. Ross, L. Pilz, B. Cabrera, F. Brachten, G. Neubaum, and S. Stieglitz, Are social bots a real threat? an agent-based model of the spiral of silence to analyse the impact of manipulative actors in social networks, *European Journal of Information Systems* **28**, 394 (2019).
 - [10] D. Takeuchi, G. Tanaka, R. Fujie, and H. Suzuki, Public opinion formation with the spiral of silence on complex social networks, *Nonlinear Theory and Its Applications*, *IEICE* **6**, 15 (2015).
 - [11] W. Waldherr and M. Bachl, Simulation gesellschaftlicher Medienwirkungsprozesse am Beispiel der Schweigespi-

- rale, in *Rezeption und Wirkung in zeitlicher Perspektive* (Nomos Verlagsgesellschaft mbH & Co. KG, 2011) pp. 235–252.
- [12] P. Gawronski, M. Nawojczyk, and K. Kulakowski, Opinion formation in an open system and the spiral of silence, *Acta Physica Polonica A* **127** (2015).
 - [13] D. Sohn, Spiral of silence in the social media era: A simulation approach to the interplay between social networks and mass media, *Communication Research* (2019).
 - [14] M. Granovetter and R. Soong, Threshold models of diversity: Chinese restaurants, residential segregation, and the spiral of silence, *Sociological Methodology* **18**, 69 (1988).
 - [15] M. A. Krassa, Social groups, selective perception, and behavioral contagion in public opinion, *Social Networks* **10**, 109 (1988).
 - [16] S. Banisch, F. Gaisbauer, and E. Olbrich, Social feedback theory: from social neuroscience to public opinion expression, forthcoming.
 - [17] S. Banisch and E. Olbrich, Opinion polarization by learning from social feedback, *The Journal of Mathematical Sociology* **43**, 76 (2018).
 - [18] A. Kianercy and A. Galstyan, Dynamics of Boltzmann Q learning in two-player two-action games, *Physical Review E* **85**, 041145 (2012).
 - [19] P. R. Neary, *Multiple-group games*, Ph.D. thesis, UC San Diego (2011).
 - [20] P. R. Neary, Competing conventions, *Games and Economic Behavior* **76**, 301 (2012).
 - [21] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, Modeling echo chambers and polarization dynamics in social networks, *Phys. Rev. Lett.* **124**, 048301 (2020).
 - [22] E. Lee, F. Karimi, C. Wagner, H.-H. Jo, M. Strohmaier, and M. Galesic, Homophily and minority-group size explain perception biases in social networks, *Nature human behaviour*, 1 (2019).
 - [23] D. Centola, J. Becker, D. Brackbill, and A. Baronchelli, Experimental evidence for tipping points in social convention, *Science* **360**, 1116 (2018).
 - [24] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*, Vol. 8 (Cambridge university press, 1994).