Steinbacher, Mitja; Steinbacher, Matjaž; Knoppe, Clemens

Article — Published Version

Opinion Dynamics with Preference Matching: How the Desire to Meet Facilitates Opinion Exchange

Computational Economics

Suggested Citation: Steinbacher, Mitja; Steinbacher, Matjaž; Knoppe, Clemens (2023) : Opinion Dynamics with Preference Matching: How the Desire to Meet Facilitates Opinion Exchange, Computational Economics, ISSN 1572-9974, Springer US, New York, Vol. 64, Iss. 2, pp. 735-768, https://doi.org/10.1007/s10614-023-10455-7

Computational Economics (2024) 64:735–768
https://doi.org/10.1007/s10614-023-10455-7

Opinion Dynamics with Preference Matching: How the Desire to Meet Facilitates Opinion Exchange

Mitja Steinbacher[1] · **Matjaž Steinbacher**[2] · **Clemens Knoppe**[3]

Accepted: 10 August 2023 / Published online: 2 September 2023

Abstract

The paper reexamines an agent-based model of opinion formation under bounded confidence with heterogeneous agents. The paper is novel in that it extends the standard model of opinion dynamics with the assumption that interacting agents share the desire to exchange opinion. In particular, the interaction between agents in the paper is modeled via a dynamic preferential-matching process wherein agents reveal their preferences to meet according to three features: coherence, opinion difference, and agents' positive sentiments towards others. Only preferred matches meet and exchange opinion. Through an extensive series of simulation treatments, it follows that the presence of sentiments, on one hand, hardens the matching process between agents, which leads to less communication. But, on the other hand, it increases the diversity in preferred matches between agents and thereby leads to a better-integrated social network structure, which reflects in a reduction of the opinion variance between agents. Moreover, at combinations of (a) high tolerance, (b) low sensitivity of agents to opinion volatility, and (c) low levels of confidence, agents are occasionally drawn away from the consensus, forming small groups that hold extreme opinions.

Keywords Opinion dynamics · Sentiments · Bounded confidence · Matching · Simulation treatments · Agent-based model · Social network

✉ Clemens Knoppe
knoppe@economics.uni-kiel.de

Mitja Steinbacher
mitja.steinbacher@kat-inst.si

Matjaž Steinbacher
matjaz.steinbacher@gmail.com

[1] Faculty of Law and Business Studies, Catholic Institute, Ljubljana, Slovenia

[2] Fund for Financing the Decommissioning of the Krško Nuclear Power Plant and Disposal of Radioactive Waste, Krško, Slovenia

[3] Institute of Economics, Kiel University, Kiel, Germany

1 Introduction

The study of opinion dynamics is a part of a domain of social interaction. Early beginnings of the opinion update theory in social science go back at least to the seminal opinion update model by DeGroot (1974). Since then the study of the opinion dynamics has attracted growing attention across a number of disciplines. In the last decade, the tractability of computer simulation and the availability of sufficient computer power have attributed to the rise of agent-based models of social interaction, in which agents have been transforming from homogeneous factors to more and more comprehensive heterogeneous actors (Macy & Willer, 2002). For instance, see an exemplary agent-based study into the collective behavior of social systems by Epstein and Axtell (1996), where authors grow economic relations from bottom up and show that even simple rules of social interaction among agents can produce stylized macro patterns. Similarly, Axelrod (1997) applied an interactive model of agents with a number of discrete features to study the dissemination of culture. The approach taken in the present paper falls in the domain of models of social interaction, based on opinion formation models with bounded confidence, as originally proposed by Hegselmann and Krause (2002, 2005); Deffuant et al. (2000, 2002).

Ultimately, the dynamics in models of social interaction depend on the way agents receive, process, and respond upon information as well as upon their interaction patterns (Macal & North, 2005; Altafini, 2013). Agents' responses to information and their interaction patterns necessarily evolve within the model. That is, the former are guided by behavioral features of agents that determine their preferences about own state or the state of their neighbors, while the latter depend on the way agents are connected among each other in their social environments and how they choose their counterparts.

However, in a current state of the literature, agents update their opinions according to own attitudes towards opinions of their counterparts. These attitudes are usually coupled to the opinion update rule and they might include features, such as stubbornness and confidence (Deffuant et al., 2000; Weisbuch et al., 2002), quorum (Ward et al., 2008), honesty (Dutta & Sen, 2012) or emotions (Schweitzer et al., 2019; Sobkowicz, 2012). This setup prevents, for instance, a study of agents' desire to meet a particular agent in isolation from the opinion update. In reality, the decision to meet with someone must be mutual for both agents taking place in the opinion exchange and it has to be made prior to their meeting.

To study the opinion update without the interference of coupling the opinion adoption to the meeting preferences, this paper sees the opinion update update as subordinated to the preference-based matching process. To achieve this aim, the paper decouples the matching of agents from the opinion update altogether. Here, note that the decoupling does not imply that the desire to meet with someone cannot not be driven by similar influences to those of the opinion update. It only means that the matching and the opinion update are two distinct processes that should stay decoupled. In our case, agents are first classified by their mutual

preferences to meet and only those agents that have been mutually paired ultimately take place in the opinion exchange. The classification of pairs proceeds in lines of a matching theory, the approach particularly popular in economics and appropriate for our task at hand. See Gale and Shapley (1962), Roth and Sotomayor (1992) and Roth (1982) for more information on basic fundamentals and the use of the matching theory in economics and social sciences. Only after this matching stage, we adopt a common framework of the opinion update as used in the literature (Deffuant et al., 2000; Hegselmann & Krause, 2002, 2005; Jadbabaie et al., 2003; Blondel et al., 2009).

In our case, the ultimate decision to meet an agent or not will be mutual and will be made beforehand by both agents. Each agent will form own lists of preferred matches in her neighborhood. These lists will form a basis for the pairwise matching of agents in preferred pairs. For this end, we will implement Irving (1985) efficient roommate matching algorithm as proposed by Sotomayor (2005), whereby the preference lists for the efficient roommate matching will be implemented as relative radial-basis scores. This turns out to be a computationally simple solution that allows the inclusion of a rich set of behavioral features in the matching process as well as any combination of their eventual co-dependencies. Radial-basis scores as used here in this paper belong to the sub-field of the radial basis kernel (RBK) classification. The RBK is a widely used and highly appreciated technique within a broad scope of the implementation of the artificial intelligence methods (Patle & Chouhan, 2013). This flexibility of the matching theory and its suitability to be dealt with computationally efficient classification techniques seems a promising way of bridging together social psychology, behavioral economics, computational social science, and artificial intelligence not only within the scope of the opinion update, but also within a wider spectrum of models of social interaction and social learning.

The preference score depends on the opinion difference, as well as human sentiment, modelled as a random variable, and neighbors' opinion coherence. In particular, the main focus of this paper is to stress the facilitating role of agents' positive sentiments of approbation towards each other, first, for the opinion exchange, and, second, for the opinion dynamics as a whole. The sentiments will be added among behavioral features that will shape the matching process between agents during the classification stage. Our motivation is straightforward. Namely, human sentiments of approbation are a long-forgotten concept in economics, despite the fact that they have been rooted in the study of interpersonal relations in economics since the inception of economics as a science (Smith & Wilson, 2019). Moreover, we add a coherence feature, so that agents only listen to those counterparts, whom they consider holding a valid opinion. Validity is defined as opinion persistence, i.e. it controls for the respect of a minimal coherence on part of agents in the opinion model. The coherence is an appealing behavioral feature that has yet to find its way into the main body of the opinion formation theory.

Results in this paper are intriguing and yet straightforward. In the presence of positive sentiments of approbation, agents meet less, as it is harder for them to find mutually preferred matches, but the matches that take place are more diverse. That is, agents are more prone to meet with various neighbours rather than always meeting with the same ones. Hence, the positive sentiments facilitate diversity in

the opinion exchange, as agents do not merely reinforce their own beliefs by meeting the same people all of the time. Consequently, the population as a whole exhibits improved global opinion convergence. It is not surprising that this convergence turns out to be more efficient in simulation treatments with a stronger presence of sentiments, where we denoted less opinion exchange and more opinion changes per meeting. And lastly, results show that a greater presence of sentiments facilitates a substantial reduction in the opinion variance while sentiments might even reduce the impact of extremely in-confident agents[1] upon others. In addition to that, the coherence feature, if strict enough, prevents extremism by regulating the impact of hyper in-confident.

The paper is organized as follows. Section 2 discusses related literature. This section intends to explain some of the main theoretical and conceptual backgrounds of the approach taken in this paper and link them to the relevant literature. The discussion in this part tries to stay informative, but it is by no means exhaustive. In particular, the paper is grounded in social sciences and uses a simulation based approach to implement its main methodological contribution, which is a separation of the opinion dynamics in two phases, preference-based matching between agents and pairwise meetings of agents. Due to an interdisciplinary construction of the paper, a bit deeper exposition of theoretical backgrounds is necessary. Section 3 presents a standard opinion model within the preference based matching framework. A general formulation is followed by the presentation of the matching between preferred pairs that is developed as an independent preference based ranking. This section keeps a mathematical disposition of the standard opinion model at the minimum, as the model is well known in the literature and does not need much further introduction. However, some additional space in this section is dedicated to the presentation of the matching mechanism, which is at the core of the approach here, particularly to the classification of preference scores by the radial basis kernel. Section 4 starts with the presentation of simulation-based treatments and proceeds with a detailed exposition of the main results. The appearance and the role of extremely (hyper) in-confident agents is discussed first and it is followed by the findings about the role of the model's parameters for the opinion update. The second part of this section is fully devoted to the presence of sentiments in the preference based matching process and how they can facilitate the opinion exchange. Finally, Sect. 5 sheds some light on the main conclusions and reveals some suggestions for future research with motivation for a further expansion of models of learning and social interaction.

2 Related Literature

2.1 Opinion Models: Background

Early contributions in opinion dynamics modelling, notably Weidlich (1971); Weidlich and Haag (1983) focused on discrete opinions, i.e. opinions $s \in [-1, 1]$, based on models from physics, such as the Ising spin model. However, the present

[1] As defined later in the sequel, extremely in-confident agents adopt changes in their opinions that are larger than overall distances to opinions they receive from their counterpart agents.

paper primarily focuses on continuous opinions along the real line. A seminal contribution by DeGroot (1974) provided analytical solutions for the convergence of opinions, i.e. consensus, in the simplest possible form. This model is mathematically equivalent to a heat diffusion process (Weisbuch et al., 2005), albeit discretized in space through the social network structure.

Subsequently, this model was expanded upon by Hegselmann and Krause (2002, 2005) ("HK model") and Deffuant et al. (2000), Deffuant et al. (2002) ("DW model"). Both models have in common that they incorporate confidence bounds, i.e. agents only change opinions upon hearing others' opinions that are not too different, defined by a suitably chosen parameter. In the HK model, agents adopt opinions of all those that are within the confidence bound (or, in networked versions, all neighbors' opinions that are sufficiently similar). In the DW model, agents interact pairwise and adopt and intermediate opinion, often the unweighted mean of both, as long as opinions are close enough to each other. These models form the basis of the model presented in this paper, as well as a vast body of research exploring additional mechanisms. The basic versions of these models have been comprehensively reviewed and analyzed by Lorenz (2007, 2010).

Another closely related research strand concerns social learning (Acemoglu & Ozdaglar, 2011), which employs similar mathematical structures to investigate the convergence of societies toward a ground truth (e.g. Golub & Jackson, 2010). Rather than assuming pure opinions, in these models it is typically assumed that a ground truth exists and convergence towards that truth through social interaction is studied. Convergence features in these contributions are typically studied mathematically, as opposed to the common simulation-based approach in opinion dynamics models.

A more comprehensive review of the evolution of opinion formation modelling, including more recent contributions, is presented by Noorazar (2020), who studies milestones of the discrete and continuous opinion models. The paper is focused on extensions of the state of the art opinion models by biased agents, stubborn agents, manipulative agents, the emergence of power, repulsive agents, and various uses of opinion models with different noise-based features. Peralta et al. (2022) provide a condensed and detailed review of the opinion dynamics models classified into models with discrete and continuous opinions. Particularly, empirical validation with data from elections and polls or experimental data is discussed.

2.2 Structures of Social Interaction in Opinion Formation Models

Opinion formation models are particularly focused on the importance of interactive structures in society that leads from simple behavioral rules to non-trivial macro-level dynamics. The following subsection reviews some contributions that study individual rules of interaction, as well as meso-scale structures, such as diverse network architectures. These models often rely on simulations, as those allow for a wider range of mechanisms that can be studied.

Examples of these lines of research include the study of the spread of extremism in society (Deffuant et al., 2002), the study of crime and social interaction (Glaeser et al., 1996), the spread of cultural traits (Büchel et al., 2014), the study of mass media and public opinions Hu and Zhu (2017), the study of social mobility (Topa & Zenou, 2015). The behavior of agents in these kinds of models is usually driven by features such as uncertainty (Deffuant et al., 2002), prejudice (Friedkin & Johnsen, 1990), opposition (Galam, 2004), influence (Acemoglu et al., 2010; Watts & Dodds, 2007).

An important feature in the literature that relates to our approach are heterogeneous confidence bounds. For instance, Lorenz (2010) studied the interaction of closed-minded (small confidence bounds) and open-minded (large confidence bounds) agents and found that, unlike in models of homogeneous bounds, there is a possibility that the consensus can drift off the center of the initial distribution. Kou et al. (2012) motivate heterogeneous confidence bounds with “complex physiological or psychological factors". Finally, Zhang et al. (2017) model time-varying confidence bounds. In our case, confidence bounds are drawn from a normal distribution, allowing for curious examples of extreme in-confidence.

Urena et al. (2019) review literature on how trust, reputation and influence propagate on on communication platforms, such as social networks. In line with the idea of trust, Duggins (2014) implements susceptibility to extreme opinions in computational experiments. A high susceptibility score implies that agents exert less influence on others, i.e. they are not trusted as much. The notion of trust and influence is embedded in our model through the coherence feature, whereby agents’ opinions are considered invalid if they fluctuate too much. Rather than reacting to the amplitude of neighbors’ opinions, agents care whether there is much informational content to others’ stances on the subject matter.

Since the seminal contribution of DeGroot (1974), opinion formation models are typically implemented on some type of network structure through which agents communicate. A variety of network structures have been applied for this purpose, such as random graphs, Watts-Strogatz small-world networks (Watts & Strogatz, 1998; Steinbacher & Steinbacher, 2019), and scale-free networks (Barabási & Albert, 1999; Das et al., 2014). As a typical prototype that replicates important characteristics of real-world social networks, specifically short distances between agents and high clustering, the underlying network in our model is also a small-world network. For instance, Pan (2012) uses probabilistic approach to study the role of standard network topologies, such as small world networks, star networks,[2] and scale-free networks, for the consensus in the opinion update. Alternatively, agents in the opinion update literature have also been placed in a social setting based on the closeness of their opinions or beliefs without the use of network topologies (Glass & Glass, 2021).

Particularly interesting is also the emerging literature on dynamic network structures, as in Wu et al. (2022); Kozma and Barrat (2008). In these models, agents can cut connections to neighbors they disagree with and replace those with new links. While we rely on a static network structure, our idea of choosing whom to meet within a given neighborhood, implemented through a preference-based matching

[2] As a name suggests, in a star network every agent is connected to a central agent.

algorithm (see the following subsection), is closely related to choosing suitable neighborhoods.

A potentially interesting extension on the use of network structures in opinion formation models are in multilayer networks. These allow the implementation of different types of connections (family, friends, colleagues), different topics etc. Diakonova et al. (2016) shows that the resulting dynamics from interaction on several layers cannot be reduced to a process on a single layer. This result has been confirmed analytically by Zhang et al. (2017). Battiston et al. (2017) applies a multilayer structure to Axelrod's model of the cultural dissemination and shows that genuinely novel behavior emerges, which allows for the existence of multiculturality despite the combined pressures of globalization and imitation.

A common shortcoming in these models is a lack of empirical validation. A step in this direction has been made by Grimm and Mengel (2020), conducting laboratory experiments to study belief formation and finding that results tend to be inconsistent with the naive learning process in the standard DeGroot (1974) model.

2.3 Preference-Based Matching

Agents in this model do not meet their neighbors at random, as they usually would in most opinion formation models, but through a separate matching process. At the core of this contribution is the idea that agents have preferences over whom to meet and act upon those. The particular matching algorithm that we implement is the roommate matching algorithm by Irving (1985). In the present paper, agents need to mutually agree to the meeting, which resembles the roommate matching problem. The approach has been particularly popularized by Roth and Sotomayor (1992) in the Handbook of Game Theory with Economic Applications. Since then, it has received many implementations in different domains of behavioral game theory (e.g. for more information on behavioral game theory, see the seminal work by Camerer, 2011) and social systems with heterogeneous and interacting agents, such as the ride-sharing matching (Wang et al., 2018), or the mentor-mentee matching on colleges (Haas et al., 2018), for instance. Closely related to the use of the matching of mutually preferred pairs of agents in the present paper, is the stable college admission and marriage problems, initiated long time ago in the works of Gale and Shapley (1962) and McVitie and Wilson (1971). Other related applications is the efficient matching of buyers and sellers in a bipartite network in Kranton and Minehart (2001), as well as job search and match problems (McCall, 1970).

The argument to include a preference-based matching process is essentially psychological. In particular, confirmation bias (Wason, 1968; Ross & Anderson, 1982) motivates this behavior of agents to not only be more likely to trust others with similar opinions, implemented by confidence bounds, but also actively seek out information that confirms their beliefs. Confirmation bias has been applied by authors such as Del Vicario et al. (2017), where agents rewire their connections in order to minimize disagreement. It is also the foundation for the emergence of echo chambers due

to the segregation into like-minded groups, as studied by Levy and Razin (2019), as well as Brugnoli et al. (2019), who find empirical evidence for it in Facebook data. On the flip side, one can interpret this mechanism as an avoidance of cognitive dissonance, that could arise through the exposure to opinions that contradict one's own, as in Li et al. (2020). In the present paper, confirmation bias is implemented through the preference score, which partially consists of the similarity in opinions.

As for empirical studies of matching, Arteaga et al. (2022) analyze how online matching platforms shape applicants' beliefs about schools. The confirmation bias has been assessed in laboratory experiments by (Zou & Xu, 2023), who show that not only similarity in opinions, but also similarity in other aspects matters. In our model, agents do not have other characteristics shaping their identities, but the idea can be subsumed under the notion of sentiments, which form part of the preference score.

2.4 Sentiments in Social Interaction

Our paper tries to enrich the social simulation model of opinion dynamics by adding a notion of a mutual desire to meet in a standard setting of opinion dynamics. In particular, it relates to the concept of mutual self-interest in which cooperative outcomes emerge endogenously. Smith and Wilson (2019) pawed a way for a reconsideration of social interaction between people in economic science. According to their setting, people remain self-interested players, but are bound within social relationships, where they exert mutual feelings of approbation or disapprobation towards each other.

Feelings of emotions have been highlighted within agent-based models of social simulation. For instance, Lejmi-Riahi et al. (2019) study emotional experience at workspace, whereby emotional state is assumed to be an important cognitive factor during work activities, such as decision-making, attention, memory, perception and learning. However, the study of emotions within the agent-based models of opinion dynamics is still in its early stage. Authors such as Bagnoli et al. (2007) have coupled the opinion update rule with the the affinity towards other agents. Similarly, Schweitzer et al. (2019) include emotions as drivers for opinion polarization. These models can be seen complimentary to the approach taken in this paper, as we completely decouple sentiments from the opinion update rule.

3 Standard Opinion Model with Preference-Based Matching

3.1 Standard Opinion Model: A General Formulation

In the standard opinion model (Deffuant et al., 2000; Hegselmann & Krause, 2002), agents update opinions at each time by incorporating some fractions of beliefs of their neighbors into their own beliefs. Let agent i have opinion $x_i(t)$ lying in some

interval $[-1, 1]$ at time t. Now, let agent i meet with agent j at time t. The opinion of agent i is then updated as follows:

$$x_i(t+1) = x_i(t) + \begin{cases} \lambda_i\big(x_j(t) - x_i(t)\big) & \text{if } d_{ij}(t) \le \theta_i \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where λ_i represents a level of confidence that agent i exerts upon the opinion of agent j, $d_{ij}(t)$ represents the absolute difference between i's and j's opinions at time t, and θ_i represents a boundary condition (usually understood in terms of agents' tolerance towards opinion differences), above which agent i never changes own opinion.

Now, let us assume agents exchange opinion in a standard setting. Further, let $x_{ji}(t_f)$ denote the opinion agent j expressed to agent i at their first meeting, and let $x_{ji}(t_l)$ be the opinion agent j expressed to agent i at their last meeting before time t. Now, let agent i be able to remember these two opinions of agent j and let it consider weighted differences between current opinion $x_j(t)$ and both past opinions of agent j as a measure of a propensity of agent j to the opinion change in the following simple manner:

$$s_{ij}(t) = \kappa\big(|x_j(t) - x_j(t_f)|\big) + (1-\kappa)\big(|x_j(t) - x_j(t_l)|\big), \;\; 0 \le \kappa \le 1, \tag{2}$$

such that the parameter κ is the relative weight for the long-term shift in opinion versus the most recent shift in opinion. Let each agent i reassess s_{ij} after each encounter with the neighboring agent j and let these scores be private information.

Say, agent i is willing to adopt changes to own opinion, only if it considers the opinion of the neighboring agent j as sufficiently stable, that is, having a sufficiently low estimated propensity to the opinion change s. Now, say agent i classifies agent j as changing opinion too much, if $s_{ij}(t)$ is above the highest acceptable level γ_i. One can easily extend the standard model with the propensity to opinion change:

$$x_i(t+1) = x_i(t) + \begin{cases} \lambda_i\big(x_j(t) - x_i(t)\big) & \text{if } d_{ij}(t) \le \theta_i, s_{ij}(t) \le \gamma_i \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The model in Eq. (3) will serve as a baseline model for the purposes of this paper.

3.2 Independent Preference-Based Ranking

In this subsection, we briefly describe the preference-based ranking for distinguishing preferred agents from non-preferred agents. As depicted in Fig. 1, the ranking is based on the radial basis kernel, where the inputs consist of different criteria that determine agent's personal preferences with respect to other agents.

By assumption, let each agent know exactly, which neighboring agents are her preferred agents for the next meeting. Let this information be learned through agents' mutual interaction and let it be private. Now, imagine there exists a unique ideal agent j^* for each agent i, such that she is always assigned a maximal possible preference score $\pi^*_{ij} = 1$ by any agent i. Now, let each agent i score her neighbors $j \in N(i)$ by some π_{ij} whereby each agent i appreciates the neighboring agents

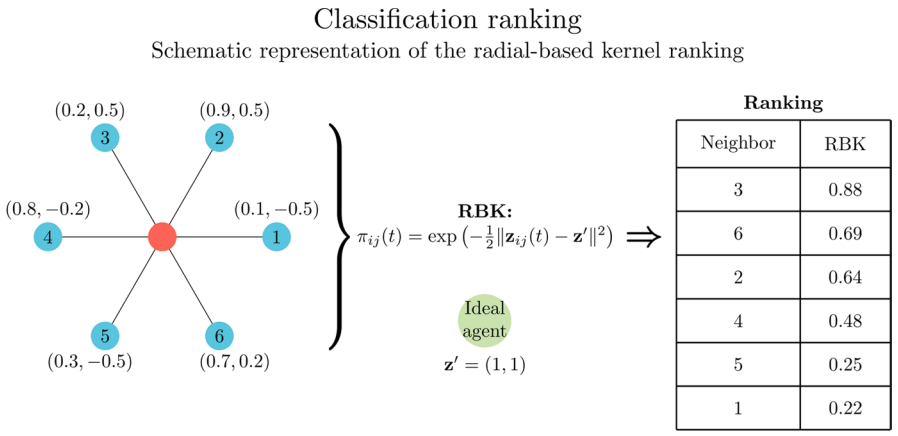

Fig. 1 Criteria and the preference-based ranking of agents: An agent (red node) observes a feature vector of all its neighbours (blue nodes) and evaluates it against the ideal vector z', using Eq. 4. The higher the score, the greater the desire to meet that neighbor, resulting in a transitive ranking. The example feature vectors have only two entries, while we are using three in the model (functions of opinion differential, opinion persistence, and sentiment) (color figure online)

j at time t by her proximity to the ideal agent, such that the larger the gap, the lower the appreciation, and vice versa.

The question of ranking neighbors by their proximity to the ideal agent is a classification problem for agents in the model. In our case, the classification is rooted in different criteria. In particular, in order to implement computationally efficient classification, we transform this multidimensional information via the radial kernel into the rankings of neighbors $\pi_{ij} \in (0, 1]$, such that the ideal agent has the score of 1. Note that the lowest score is zero at infinity, that is, the least ideal agent is infinitely less appreciated than the ideal agent, accordingly. We implement agents' mutual scoring via the following radial basis kernel:

$$\pi_{ij}(t) = \exp\left(-\frac{1}{2}\|\mathbf{z}_{ij}(t) - \mathbf{z}^*\|^2\right), \tag{4}$$

such that $\mathbf{z}_{ij}(t)$ is a feature vector of agent's i personal preferences towards agent j. In principle, the feature vector may consist of any quantifiable feature that can be expressed as a score on the [0, 1] interval. Moreover, the feature vector can include multiples of different features (i.e. to account for the presence of dispersion) or their cross-products (i.e. to account for the co-variation of features).

Preference-based ranking in this paper will include a set of the following basic behavioral features that each agent can easily assess: agreeableness in the opinion (i.e. an opposite to the opinion difference), coherence in the opinion (i.e. opposite to the propensity to the opinion change defined in Eq. (2)), and sentiment noise ω, defined as a random noise $\omega \sim \mathcal{U}(0, \alpha)$ to capture the presence of approbation felt by

a particular agent towards other agents in the neighbourhood. Hence, let each agent i assess a feature vector $\mathbf{z}_{ij}(t)$ of her neighbour j at time t as follows

$$\mathbf{z}_{ij}(t) = \left\{ 1 - \frac{1}{2}|d_{ij}(t)|, 1 - \frac{1}{2}s_{ij}(t), \omega_{ij} \sim \mathcal{U}(0, \alpha) \right\}, \tag{5}$$

such that $|d_{ij}(t)|$ is the absolute difference in opinions of agents i and j, $s_{ij}(t)$ is a propensity of agent j to the opinion change as defined in Eq. (2), and ω_{ij} is an independent sentiment-effect (uniform at random) that expresses a feeling of approbation felt by agent i towards agent j. Note that the feeling of approbation might be modelled as a process on its own within the opinion model, but this would depart us from the main focus of this paper that is to understand different ways in which changes in sentiments might facilitate the opinion exchange and ultimately affect the opinion dynamics via the matching process that is decoupled from the opinion update.

3.3 The Matching of Preferred Pairs

Central to the present model is that agents do not meet at random, but create preference lists of their neighbors and always try to meet with the most preferred one. However, agents do not simply choose to meet with the highest-ranked neighbor. The preference to meet must be mutually ensured, as highest ranked neighbor of any agent might prefer someone else, not this particular agent on whose list it ranks highest. We face a matching issue here, such that enables a pairwise pairing of agents, where appropriate agents will ultimately meet based upon their mutual rankings. In order to solve this problem, we apply the Roommate Matching algorithm as proposed by Irving (1985). This algorithm ensures stable matches, whereby every agent meets the highest-ranked neighbor, given this neighbors' preferences. In particular, a stable matching (ij) is a matching, where no two agents i, j are left unmatched with each other, if they prefer each other over their matches k, l (Roth, 1982). Rigorous treatments, including formal proofs, of the algorithm can be found in many places (E.g., such as Gale & Shapley, 1962; Irving, 1985; Roth, 1982; Roth & Sotomayor, 1992). Due to the importance of the algorithm to the model, a brief explanation of the main mechanisms of the algorithm is in place.

Starting with the agents' rankings, the algorithm can be split into two phases: in phase 1, each agent i proposes to the highest ranked neighbor j on her list. The agent that has been proposed to tentatively accepts and eliminates all agents from her list that are lower in her own preference rankings. Note that any deletion of agent j in agent i's list is matched by agent j deleting agent i from her own list, so that no proposals can be made to neighbors that have, implicitly or explicitly, rejected her already. The deletion is not going to have negative consequences for j, since by design of the algorithm, she cannot get a worse matching than i in anyway.

If agent j receives a second proposal from an agent k that is higher-ranked than the initial proposal by agent i, the agent j has to eliminate agent i from her list (and all other agents that are ranked lower than k). As the deletion is again mirrored by the agent i, the i is now without a match and gets to propose to the next highest on

her list. Once each agent has made a proposal that has not yet been rejected, or has run out of neighbors to propose to, the first phase is terminated.

Hence, at the end of phase 1, each agent has either a tentative confirmation by another or has been rejected by her entire neighborhood. If at this stage, any agent i has only one agent j left in her reduced list, it immediately follows, that this also holds true the other way around, so that the two agents form a stable match. They do not require to be processed any further in phase 2. Agents that have offers from more than one neighbor, i.e. their reduced lists contain several neighbors, are then treated in the second phase.

In the second phase, the algorithm detects cycles in preferences and uses them to reduce the lists further. The cycles consist of two steps. The cycling phase starts from any agent, say p_i, and selects the next one, say q_i, that has to be the second in the p_i's preference list. It proceeds by selecting the next p_{i+1} agent that has to be the lowest-ranked neighbor in q_i's preference list, and so forth, until agent p_i reappears again as p_{i+k}, this time having its worst ranked neighbor attached to it.[3] Once this particular kind of cycle has been detected, all agents q_i in the cycle delete their lowest-ranked neighbors (i.e. p_{i+1}) from their preference lists. This procedure continues until all preference lists are either of length 0 or 1. See the Appendix 1 for a graphical description of this process together with some examples,

Note that the algorithm does not ensure every agent to find a match in every iteration of a particular simulation treatment. A trivial case represents an odd number of agents, where at least one agent always gets unmatched per iteration, or in specific network configurations that prevent social interaction by construction, such as in a star network, for instance. However, even in complete networks with an even number of nodes, complex configurations of network topology agent's preferences to meet each other might evolve that do not warrant agents to succeed in finding their preferred matches. However, as shown in Sect. 4, this kind of complexity in behavior indeed characterizes the matching process in our case. Depending on some of the model's parameters as well as on the presence or absence of sentiments, the number of matches per iteration can differ significantly, giving rise to differences in the opinion dynamics.

4 Results

4.1 Simulation Treatments: Setup

This section brings a condensed presentation of all parameters of the model as they were used in simulation treatments with their brief descriptions and initial values. In particular, Table 1 describes the distributions of parameters, while Table 2 complements these distributions with concrete values as they were used in simulation treatments. Two versions of the standard opinion update model with preference-based matching were tested:

[3] Note that the cycle does not necessarily start from agent p_0, but could start at any other agent p_k. In that case, agents $p_{0,1,\ldots,k-1}$ are called the "tail" of the cycle. Agents $q_{0,1,\ldots,k-2}$ in the tail do not delete any agents in their preference lists.

Table 1 Parameter values for simulation-based treatments: general descriptions

Parameter	Description	Value
N	A total number of agents in the model	300
$\langle k \rangle$	Expected number of neighbors per each agent	8
p	A share of distant connections per each agent in a network, i.e rewiring probability in the small-world model (Watts & Strogatz, 1998)	0.15
$\mathcal{G} = (N, \langle k \rangle, p)$	Small-world network	$N = 300, \langle k \rangle = 8, p = 0.15$
$x_i(t)$	Opinion of agent i at some moment in time t	$x_i(0) \in \mathcal{N}(\mu(x) = 0, \sigma(x) = 0.5)$
γ_i	A lower bound on the persistence score, below which agent i considers others' opinions as invalid	$\gamma_i \in \mathcal{N}(\mu(\gamma), \sigma(\gamma))$
κ_i	A geometric weight for the long-term versus the short term persistence. The weight κ stresses the importance as seen by agent i of the distant opinion of other agents relative to the latest opinion $(1 - \kappa_i)$ of other agents	$\kappa_i \in \mathcal{N}(\mu(\kappa), \sigma(\kappa))$
θ_i	An upper level of tolerance upon differences in opinions that is still acceptable to agent i	$\theta_i \in \mathcal{N}(\mu(\theta), \sigma(\theta))$
λ_i	A bounded confidence parameter that agent i places on agent j at changing own opinion	$\lambda_i \in \mathcal{N}(\mu(\lambda), \sigma(\lambda))$
ω_{ij}	A noisy sentiment effect (feeling of approbation) in agent i after the latest opinion exchange with agent j	$\omega_{ij} \in \mathcal{U}(0, \alpha)$
α	An upper bound on the sentiment noise ω. Note: α stays the same throughout a particular treatment	$\alpha \in [0, 1]$

1. The matching process goes on without coherence and agents do not validate opinion of their counterparts upon the meeting and
2. The matching process includes all three behavioral features including coherence and agents validate the opinion of their counterparts upon the meeting.

Both versions were separately tested in an environment with and without the presence of sentiments. In both versions, some agents appeared with $\lambda > 1$. These agents will be called hyper in-confident. As we will see, in the second version of the model, hyper in-confident agents were neutralized particularly in those treatments, where agents were in general highly tolerant towards each other. In these treatments hyper in-confident agents were considered to be incoherent and other agents simply ignored them. The first setup is a baseline scenario, while the second setup enables a study of the role that sentiments might have as facilitators of the opinion exchange also in circumstances when agents possess also some rational guidance in the formation of their preferred matching lists and some ability to asses coherence in the opinion of their counterparts upon the meeting.

The model is implemented in three main steps: first, agents form preferences over their neighbors and the matching algorithm determines the pairings. Then, agents meet, exchange opinions and update their opinions if the necessary conditions are

Table 2 Parameter values for simulation-based treatments: a general setup

Parameter	Description	Value
$\mu(\gamma)$	Population mean of γ	$\mu(\gamma) = \{0.1, 0.2, \dots, 0.9\}$
$\sigma(\gamma)$	Population standard error of γ	$\sigma(\gamma) = 0.1 * \mu(\gamma)$
$\mu(\kappa)$	Population mean of κ	$\mu(\kappa) = \{0.1, 0.2, \dots, 0.9\}$
$\sigma(\kappa)$	Population standard error of κ	$\sigma(\kappa) = 0.1 * \mu(\kappa)$
$\mu(\theta)$	Population mean of θ	$\mu(\theta) = \{0.1, 0.2, \dots, 0.9\}$
$\sigma(\theta)$	Population standard error of θ	$\sigma(\theta) = 0.1 * \mu(\theta)$
$\mu(\lambda)$	Population mean of λ	$\mu(\lambda) = \{0.05, 0.1, 0.15, 0.2, \dots, 0.95\}$
$\sigma(\lambda)$	Population standard error of λ	$\sigma(\lambda) = 0.1 * \mu(\lambda)$
α	Upper bound on sentiment effect ω	$\alpha = \{0, 0.05, 0.1, \dots, 1\}$
Initialization of the matching parameters (effective at $0 \leq t < 5$)		
$s_{ij,0\leq t<5}$	Initial persistence score by agent i towards agent j	$\gamma_{i,0\leq t<5} \sim U(0,1)$
$\omega_{ij,0\leq t<5}$	Initial sentiments felt by agent i towards agent j	$\omega_{i,0\leq t<5} \sim U(0,1)$

met (Eq. 3). Finally, agents update their preference scores, in particular the coherence score and the opinion difference, which depend on the last observed opinion of the neighbor. A completion of these three steps forms one iteration of the algorithm. A more detailed presentation is provided in form of a pseudo-algorithm in the Appendix 1. We simulate the model with 300 agents, situated on a Small-World Network, such that each agent has an expected number of eight connections (see Table 1 for details on the set-up). Here, a single complete treatment means 1, 000 iterations of the algorithm under a single parametrization of the opinion model. Altogether, in the second version of the model 9^3 different expected values of γ, κ, θ were used within 19 different expected values of λ and within 21 different values of upper bounds α on the sentiment effect ω. This gives us a maximum overall number of 290, 871 independent simulation treatments that were implemented on the second version of the model. Note that the completion of all treatments within the first version includes a ninth of the overall number of independent simulation treatments needed for the second version. Note also that each simulation treatment starts with the same network topology and that each agent in each independent treatment is given an independent set of the agent-based parameters obtained from the distributions as provided in Tables 1 and 2.

4.2 Minimum Variance and Hyper in-Confident Agents

As for the conventional parameters of the standard opinion model, our results in this paper are in line with those from the baseline model in Steinbacher and Steinbacher (2019), i.e. at higher values of the mean tolerance parameter $\mu(\theta)$ we observe a lower variance in opinion. Here, lower variance in opinion is usually a sign of convergent dynamics in the opinion distribution. This implies that as long as sufficiently diverse agents participate in

the opinion exchange, the convergent features of the standard opinion update model are at play, driving opinions closer together and reducing variance in opinion thereby.

Moreover, given the results of both versions with and without sentiments in the matching process, presented in Fig. 2, the minimal variance in the final opinion distribution appears at around $\mu(\lambda) = 0.5$, which represents the well studied point of the strongest convergence in opinions. However, note that the turning point in the opinion variance is not necessarily exactly at $\mu(\lambda) = 0.5$, as the behavioral features of agents in the model are parameterised from symmetric distributions around given expected values. Moreover, in the presence of coherence (i.e. in the second variant), minimal variance appears to happen at slightly lower values of $\mu(\lambda)$. This may be explained by the fact that λ determines the size of opinion adjustments and the demand for coherence in the opinion of counterpart agents punishes too large movements in their opinions, i.e. this might reduce volatility. In other words, a particular independent set of parameters might be obtained, such that parameters are more favorable to lower variance than parameters in some other set of parameters from these same distributions, thus opinions of agents are not necessarily the least dispersed when $\mu(\lambda = 0.5)$. In addition, treatments with sentiments in the matching process (panel on the right in Fig. 2) show substantially lower variance than treatments without sentiments in the matching process. This finding corroborates our expected account of the facilitating role of the sentiments for the opinion exchange, a conclusion, to which we will gradually arrive in the sequel. In addition to overall lower variances when sentiments are present, we can observe that the minimum is achieved at a larger range of values of $\mu(\lambda)$.

Note an increase in variance at values of $\mu(\lambda) \geq 0.9$. Such an increase is likely due to the appearance of some hyper in-confident agents i with $\lambda_i > 1$. Their appearance is expected at sufficiently high values of $\mu(\lambda)$; see Table 1 for a general description of parameters that were used in the simulation treatments and Table 2 for their values. Hyper in-confident agents adopt changes in their own opinions which are larger in value than the absolute difference between their own opinions and opinions of the agents they were matched with. The presence of hyper in-confident agents in simulation treatments can lead to the occurrence of extreme groups, particularly if these agents get contacted by agents at the tail of the opinion distribution and get dragged away from the center. This increase in variance is facilitated by tolerance of agents, i.e. higher values of $\mu(\theta)$. While it enables convergence in the presence of "normal" levels of confidence, tolerance also allows for extreme groups on the fringes when hyper in-confident agents are present.

As we can see from the bottom panel of Fig. 2, referring to the treatments of the second version of the model, which includes coherence in matching and opinion updates, the impact of fringe groups seems to disappear. We do not observe the phase transition in the variance of opinions at a combination of high values of $\mu(\lambda)$ and $\mu(\theta)$, but rather a continuous, well-shaped behavior of the opinion variance throughout the entire range of parameters. Variance is persistently lower the higher the tolerance parameter θ is. The absence of coherence scoring supports the impact of hyper-in-confident agents in this case. We will elaborate on this finding in a bit more detail in the Sect. 4.3.

Figure 3 shows one such treatment, where extreme opinions are formed: a small group of agents appears in the left corner at opinion values of -4 in the plot on the

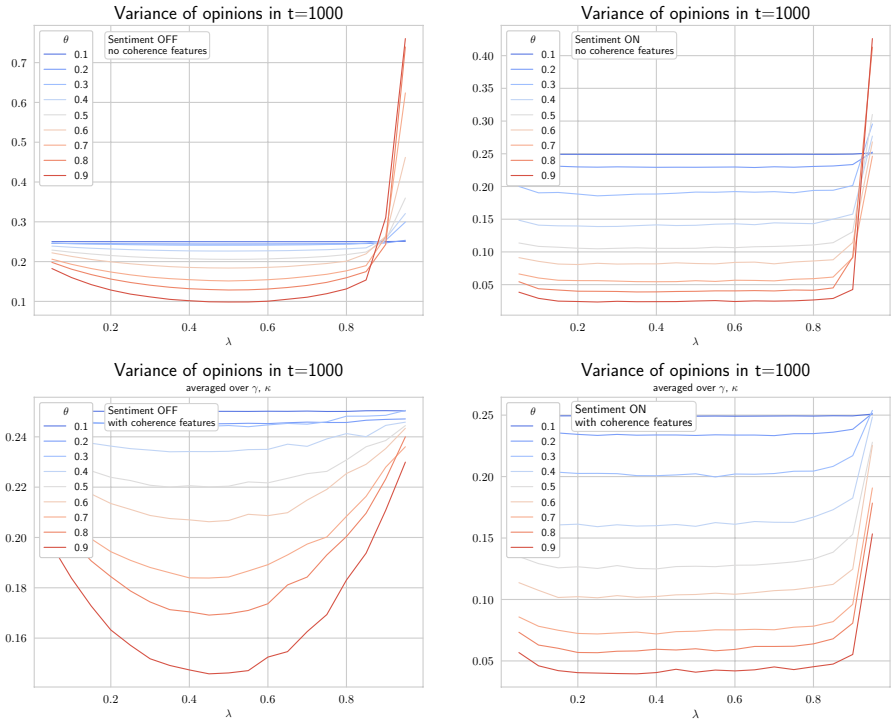

Fig. 2 Variance in opinion as a function of confidence; Left column: simulation treatments without sentiments (Sentiment OFF, $\alpha = 0$); Right column: simulation treatments with sentiments (Sentiment ON, $\alpha = 1$); Top row: simulations without coherence in either matching or opinion formation (averaged over independent simulation runs); Bottom row: simulations with coherence (averaged over values of γ and κ). Note how punishing volatility in opinions avoids the extreme increase in variance for large values of $\mu(\lambda)$

right hand side of the Figure. This is the treatment with the highest sentiment bound $\alpha = 1$ supported by a sufficiently high level of average tolerance $\mu(\theta) = 0.9$ among agents. Namely, agents with higher tolerance are more likely to be affected in the sense that they easily readjust their own opinions.

However, as we will see in the latter part of this section, after first studying the role of model parameters for the opinion exchange (Sect. 4.3), the stronger the presence of sentiments in simulation treatments the more agents tolerate diversity in opinion (Sect. 4.4). Agents with these preferences show large overall number of opinion changes and promote the convergent forces of the standard opinion model. Despite the presence of sentiments, an extreme group appeared in treatments without coherence in matching and opinion adjustment (Fig. 3). This was sufficient to substantially increase the variance of opinions, but the group remained isolated and the convergent features of the model prevailed in leading the society towards a consensus. Moreover, extreme groups only appear in few treatments, and the dispersion of the variance measure among treatments with high values of $\mu(\lambda)$ was high.

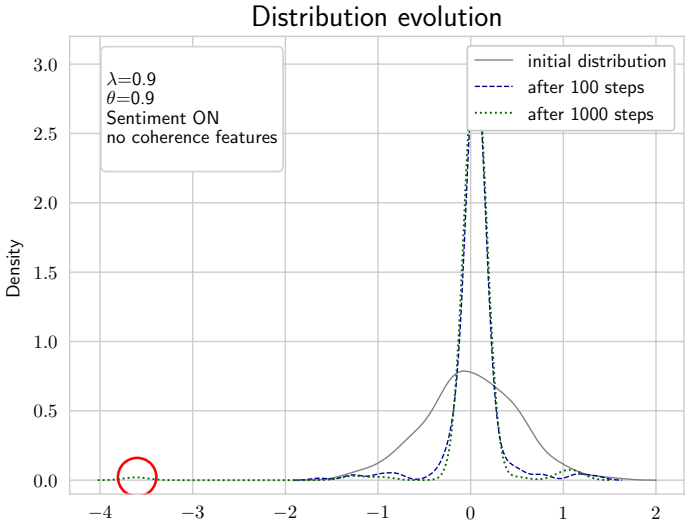

Fig. 3 Distribution of opinions: the appearance of extremes (2 agents in this case). Simulation without coherence features in matching or opinion adjustment. It seems that convergence has been reached early on ($t < 100$), as is common in models without matching (e.g. Steinbacher & Steinbacher, 2019). While we have not studied the question in detail, it indicates that asymptotic convergence is achieved

In addition, we should also stress that in treatments with coherence in the matching process and opinion validation in the opinion exchange, we might expect some dependence of the opinion dynamics upon the positioning of agents in their neighborhoods. Some agents might get stuck in their social environment, such that their neighbors label them as holding an incoherent opinion early enough in the treatment. Such agents might never get involved in the opinion exchange with others. Anyway, we have not studied the role of social environment and network effects in this paper.

4.3 The Role of the Model's Parameters for the Opinion Exchange

In what follows in this part are treatments to study the role of the main parameters in both versions of the standard opinion model with matching and coherence discussed earlier [i.e. see Eqs. (1) and (3]. In this section, we disentangle the effect of the model's parameters for the presence of convergent forces in the opinion exchange, as captured by the opinion variance. Results are shown in Fig. 4.

In general, it can be assessed, that as the coherence requirements become stricter (i.e. at lower acceptance boundaries of perceived volatility of others' opinion $\mu(\gamma)$), more and more opinions are considered invalid, which limits the convergence of opinions (top left panel). Convergence is driven by opinion changes on the micro-level, which are not tolerated in these set-ups, and hence opinions remain relatively widely dispersed.

However, we detect a phase transition as the adjustment parameter $\mu(\lambda)$ gets large: at low values of $\mu(\gamma)$, i.e. around 0.3, the role of γ of reducing the opinion variance, which we observed at lower values of λ, reverses. We now observe a sharp increase in the dispersion of agents' opinions as $\mu(\gamma)$ increases, i.e. as agents care less about the validity of opinions they encounter. The effect can also be observed in the top-right panel in Fig. 4: while a regular cyclical pattern occurs throughout the plots at $\mu(\gamma) <= 0.8$, this pattern breaks down at $\mu(\gamma) = 0.9$. At large values of $\mu(\lambda)$, i.e. in the presence of hyper in-confident agents, the variance of opinions sharply increases.

The constraint on the size of opinion adjustments is not binding when γ is large. Hence, at large values of λ, which is the size of opinion adjustments relative to opinion differences, hyper in-confident agents are not effectively regulated by their social environment and they are able to build groups of extreme opinions away from the general consensus.

The bottom panel of plots in Fig. 4 reveals the effects of κ and γ on the opinion variance and the opinion exchange. We can denote symmetric work of κ on the variance, which is expected as the parameter measures geometric weights of distant versus recent differences in opinion in the coherence score. The less important are both differences for agents, the lower the variance in opinion, and vice versa, i.e. the more important is either of these differences to agents, the larger the variance in opinion. At $\kappa = 0.5$, agents are indifferent between distant and recent changes in opinion. When agents put more emphasis on longer-term persistence, i.e. when $\mu(\kappa)$ is large, γ binds the accumulated opinion adjustments towards the center, rather than individual opinion changes. Hence, agents that start in the tails of the distribution are limited in their ability to converge towards the center. On the other hand, when $\mu(\kappa)$ is small, i.e. agents emphasize short-term persistence, any substantial change in opinion adjustment invalidates the opinion of the agent and prevents the matches from adjusting theirs. At intermediate values of $\kappa \approx 0.5$, agents allow for individual changes, as well as accumulations thereof. Hence, fewer agents' opinions are invalidated, which supports gradual convergence of opinions, as observed through lower variance.

Having said so, the bottom-right plot is more interesting. Remember, the variance is approximately the same at lower and at higher end of the κ (i.e. in treatments with and without sentiments). The more agents care about long-term changes in opinions, i.e. at larger values of $\mu(\kappa)$, the more relevant the value of γ becomes. At low values of κ, agents care mostly about recent changes in opinion, which do not accumulate over time as agents converge towards the centre of the opinion distribution. At larger values however, the accumulated differences begin to matter and create binding boundaries on the adjustment of opinions, reducing the number of opinion changes over the duration of the simulation.

Ultimately this boils down to the fact that the existence of memory on the part of agents is a precondition for the concept of coherence. However, we observe that it is not only the existence of memory as such that matters, but also how agents remember, as defined in our case by the Eq. 2, through the dependence on the value of κ. Coherence becomes more binding, and relevant to the willingness of agents

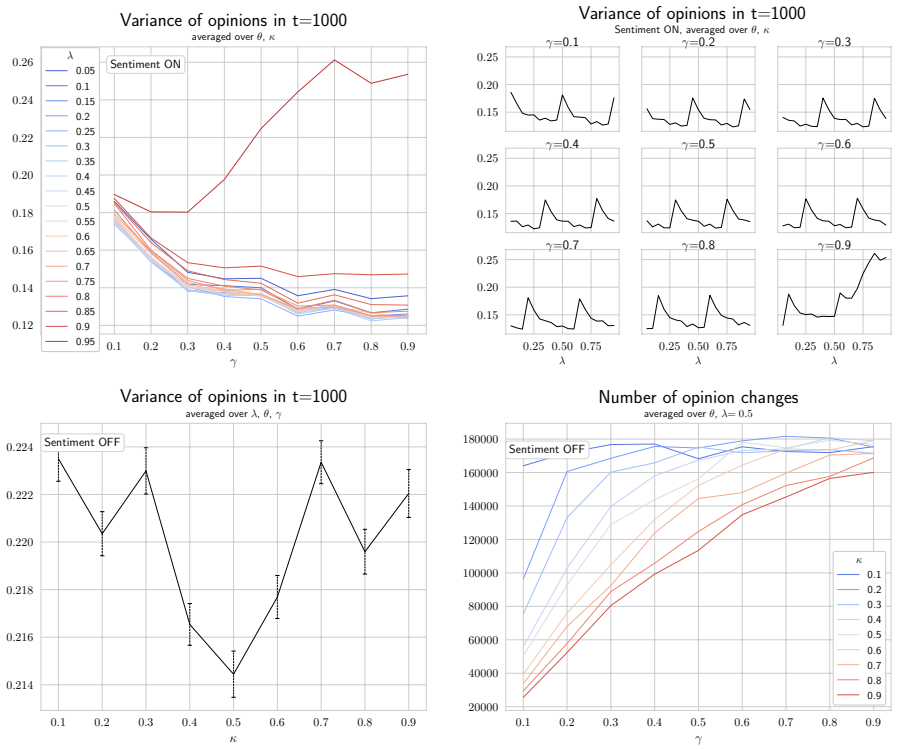

Fig. 4 Top panel: variance of opinions as functions of λ and γ. Bottom panel (left): opinion variance as a function of κ; Bottom panel (right): the total number of opinion changes as a function of γ and κ. Sentiment OFF (ON) refers to sentiment bounds $\alpha = 0$ (1)

to reconsider their opinions, as long-term memory (larger κ) gains importance and therefore opinion changes accumulate.

And lastly, a striking result that actually comes "for-free" alongside this study, is the variance reducing feature of the coherence parameter γ. Recall for a moment a sharp increase in the dispersion of opinion in treatments without coherence in the top panel of Fig. 2. Namely, what happens here is as follows: once agents are allowed to observe changes in the opinion of others and ignore agents with high propensity to change their opinions, they effectively reduce the dispersion in the opinion across the whole society. This effect can be seen by comparing the upper panel to the lower panel in Fig. 2. Agents in the second setting of the opinion model with coherence can neutralize the impact of hyper in-confident agents in the society, effectively preventing extremism.

4.4 Introducing Sentiments: How do they Facilitate the Opinion Exchange?

Finally, let us now turn our attention to the presence of sentiments and study their role as facilitators of the opinion exchange. Note that sentiments in our case are

independently determined for each agent as a measure of approbation towards other agents. Sentiments enter each agent's preference score (i.e. see the Eq. 4) as a part of the matching that takes place before the meetings between agents are implemented. In what follows, we present the main results of a rich set of treatments obtained at various levels of the upper bound α on sentiments.

In particular, let us first focus on the number of meetings in treatments with and without coherence features. According to the results in the top panel of Fig. 5, the presence of sentiments in the matching process reduces the number of meetings in both treatments with, and without coherence. This happens, because the presence of sentiments broadens the pool of potential preferred matches that can be made. Given that sentiments are independent and non-symmetric between pairs of agents, this complicates the matching between preferred pairs. As a result, they are expected to be more willing to listen to agents that hold more diverse opinions than in the case where the matching of agents depends only upon the opinion-based criteria, such as the absolute opinion difference or coherence.

However, this is only one part of the story. Let us now see, i.e. for the same setting, how sentiments interplay with the number of opinion changes. According to the bottom panel of Fig. 5, in the presence of sentiments, agents show larger willingness to adopt changes in their opinion than in treatments without sentiments, even though sentiments play no role in the decisions to adopt opinion changes. In addition, note the trajectory of a zero sentiment curve and compare it to the trajectories of nonzero sentiment curves. What we can observe here, is a comparison of two different behaviors. In the zero sentiment case, the number of opinion changes follows the level of tolerance $\mu(\theta)$ in a predictable and well-known way, i.e. the larger the level of tolerance in society, the more agents are willing to change their own opinions. On the other hand, according to our simulation treatments, even a weak presence of sentiments is sufficient to mix preference lists of agents so much that it increases the diversity of mutual pairings between agents, which can dramatically increase the number of opinion changes, particularly at the lowest to medium levels of tolerance $\mu(\theta)$.

This finding permeates through the results of our investigation into sentiment effects: upon adding sentiments into the matching process, we observe a change in the qualitative behavior of the model. This qualitative change appears in Fig. 5 as a discontinuous drop in the number of meetings and a difference in the relationship between θ and the number of opinion changes between sentiment bounds 0 and 0.05. Further increases of the sentiment bound α however only create quantitative changes, as we will discuss in detail below.

In addition, we can also detect the effect of θ and its interplay with sentiments. At $\mu(\theta) > 0.7$ (without coherence) and $\mu(\theta) > 0.8$ (model with coherence), the effect of sentiments on bringing diverse people together dissipates as agents at high enough levels of tolerance might start getting more alike faster. The tendency to mix them with diverse agents in such setting can be expected to have negligible effect for an increase in the number of opinion changes. At some point, agents are so much alike that sentiments just play no role anymore. Reinforcing beliefs by meeting similar minded agents in the no-sentiment case starts dominating in terms of the overall

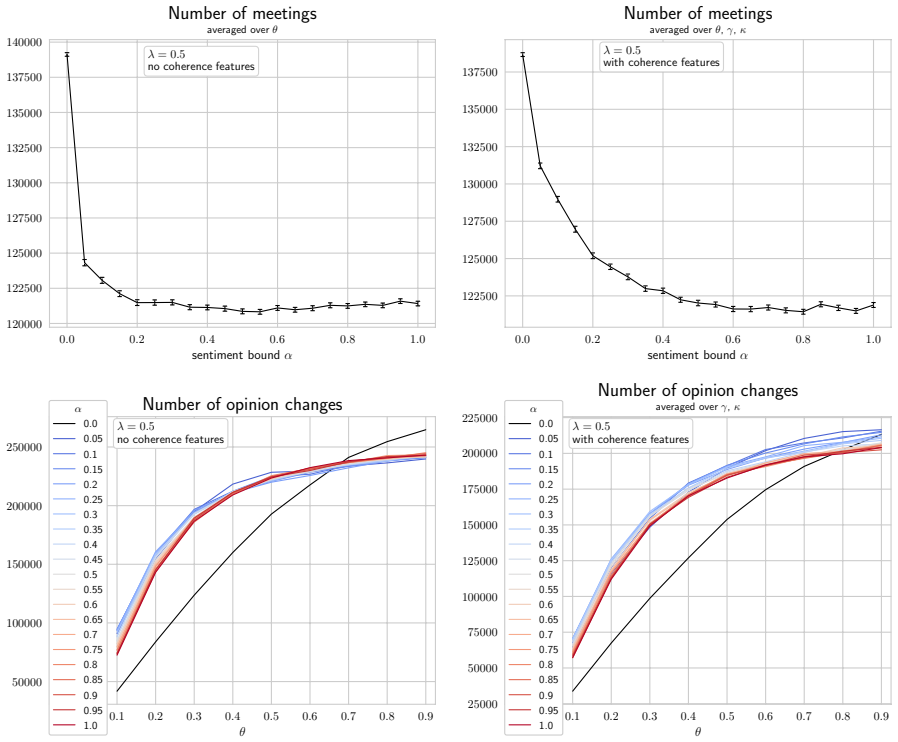

Fig. 5 Effects of sentiments on numbers of meetings and opinion changes. Top: Number of meetings; bottom: number of opinion changes; left: model without coherence features in matching and opinion adjustment; right: model with coherence features

number of opinion changes over the tendency of agents to see the diversity in opinion (i.e. at higher sentiment bounds α).

Now, let us combine our views of meeting numbers with numbers in opinion changes at Fig. 5 and think for a while of the meeting efficiency shown in Fig. 6. Again we observe the above-mentioned qualitative difference between the zero-sentiment and non-zero sentiment treatments. By adding any level of sentiments into the matching process, the ratio of opinion changes to total meetings increases sharply compared to zero sentiments. Upon increasing sentiments, this ratio continues to rise gradually until $\alpha \approx 0.3$, after which it decreases. At high values of α, agents that are too different in opinions are matched and therefore cannot exchange opinions as frequently as they do at medium levels of α. Moreover, the meeting efficiency is lower in the presence of coherence features at all levels of sentiment bound α, due to the added constraint on opinion updates.

Figure 7 shows the relative increase of meeting efficiency between sentiment bounds 0.05 and 0.95 (left panel), and 0 and 1 (right panel) over the tolerance parameter θ, respectively. Note that comparing of two non-zero levels of α yields a qualitatively different result than comparing zero sentiments with a non-zero level. The qualitative

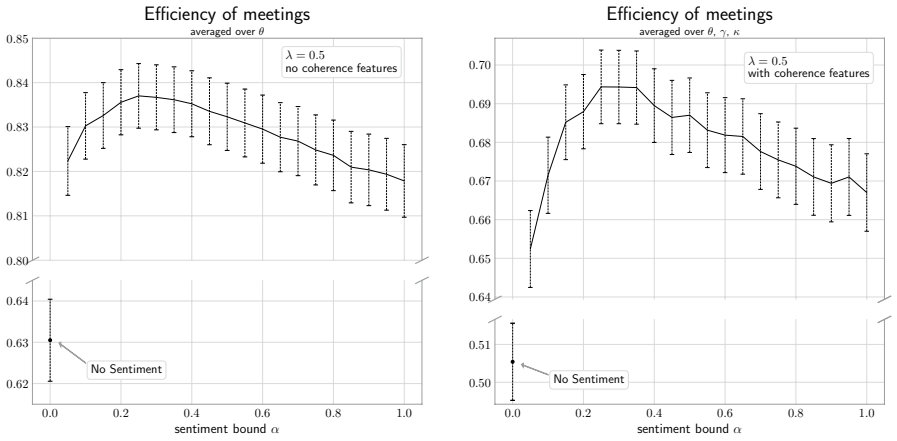

Fig. 6 Meeting efficiency, defined as $\frac{1}{2}\frac{\text{opinion changes}}{\text{meetings}}$, for models with and without coherence features in matching and opinion adjustment

difference lies in the difference of comparing levels of sentiments, given they are present, and analyzing the presence of sentiments. However, in both cases the effects of sentiments dissipate as θ increases. On the one hand, in comparing the treatments when $\alpha = 0$ with treatments when $\alpha = 1$, we can observe a monotonous relationship. The presence of sentiments matters less the more tolerant agents are.

Comparing non-zero levels of the sentiment bound α (left panel) shows a different interaction between tolerance θ and sentiments. At very low levels of tolerance $\theta = 0.1$, the increasing sentiments leads to a reduction in meeting efficiency. The agents are too intolerant to accept the added diversity of opinions they are exposed to in their meetings. However, as tolerance increases, we detect an increase in meeting efficiency as a result of increasing the sentiment bound. Hence, low levels of tolerance can be mitigated by sufficient levels of sentiment and a willingness to adopt new opinions upon meeting more diverse agents. The effect approaches 0 as $\theta \geq 0.5$. At this level of tolerance, higher levels of sentiment do not add to the meeting efficiency. This stands in contrast to the right plot, which shows the increase in meeting efficiency upon the presence of sentiments: for all levels of $\mu(\theta)$, the efficiency of meetings is substantially higher.

During our discussions so far, we have pointed to the role of sentiments to increase the diversity of meetings. Namely, with the addition of a random variable in the matching process, agents are more willing to meet agents that are not necessarily the most similar in the opinions they hold. On the other hand, without sentiments, they would prefer to meet the most similar agents only, and the preferences would be mutual between pairs of agents. The exchange of opinions would reinforce those preferences as they become more similar in opinion. Upon adding sentiments to the matching process, we can identify an increased diversity of meetings between agents. Figure 8 presents a comparison of the number of unique pairs of agents that have met, i.e. the density of the network of effective interactions, with the total number of edges in the network. The largest increase happens immediately upon adding sentiments to the model, but it increases

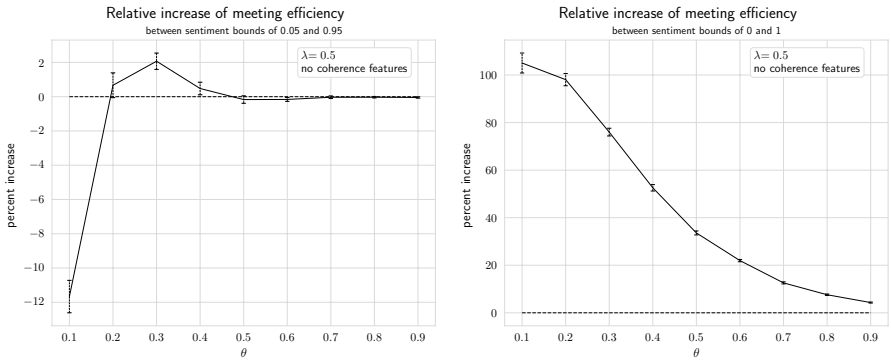

Fig. 7 The relative increase between very high and very low levels of sentiment, as regulated by θ. Left: difference between sentiment bounds 0.05 and 0.95; right: difference 0 and 1

further as the sentiment bound α increases. Sentiments ensure a well-integrated and connected society, whereas in their absence, the matching process stabilizes and always connects the same pairs. This effect is observed in both versions of the model, with and without coherence features.

Last, but not least, after we have identified the facilitating role of the sentiments for the efficiency in the opinion exchange as they allow more diversity in mutual pairings of agents before their meetings. Let us take a look at the dispersion in the opinion among agents at the presence of sentiments in the matching process. The evolution of the variance of opinions is shown in Fig. 9.

The trajectory of the variance is strikingly clear: a change in variance as a function of varying sentiment bounds α follows a continuous and monotonous relationship, except for the jump from zero sentiment case to the first non-zero sentiment case at $\alpha = 0.05$. The variance curve shows that the convergent forces increase when agents, due to the impact of positive sentiments, prefer to meet diverse agents.

Furthermore, on a side note, the reduction in variance is also driven by θ, as shown in Fig. 10. Without a sufficient level of tolerance among agents, positive sentiments would only contribute to opinion exchange, but not also to the reduction of dispersion in the opinion within the society. Hence, if agents are too intolerant too change their opinions, in almost every meeting, a partial randomization of the matching process (i.e. due to the presence of sentiments), is not able to improve consensus-finding. In particular, at $\theta \in [0.1, 0.2]$, sentiments fail to improve opinion convergence. Only at sufficiently high tolerance levels (i.e. at $\theta \geq 0.3$), the presence of sentiments reduces the dispersion in opinion, with the highest effect being denoted at medium levels of tolerance. In line with these results, the improvement due to the presence of sentiments is less pronounced at high values of θ, as agents are highly tolerant and opinions might also converge in the absence of sentiments, given some other conditions are met (i.e. such as the absence of hyper in-confident agents).

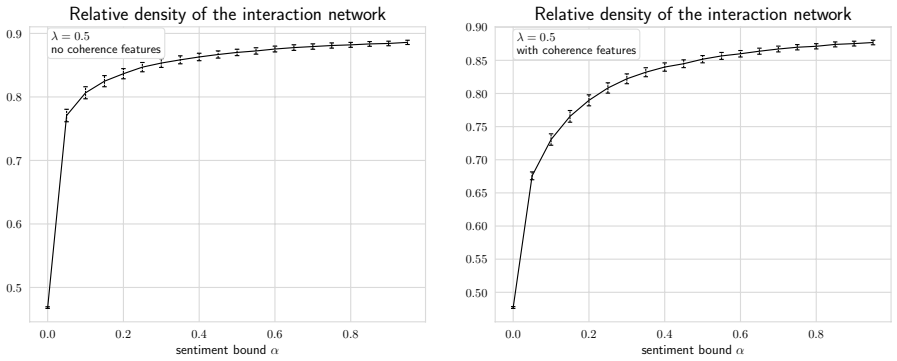

Fig. 8 The relative density of the interaction networks. Unweighted edges are drawn between agents that have been matched at any point in a simulation run. The total number of links (density of the network) is then compared to to the density of the original network that defines neighborhoods of the agents. Left: model without coherence features in matching and opinion updates; right: model with coherence features

5 Discussion and Perspectives

We have disentangled the opinion model and split it into two parts: the preference-based matching of agents and the opinion exchange. This enabled us to see behind the process of choosing counterpart agents before the opinion exchange takes place.

Some of the most standard convergent features from the opinion dynamics literature were also underlined in this paper. For a more detailed discussion about mathematical foundations of convergence in the standard opinion model, see Acemoglu and Ozdaglar (2011) and Acemoglu et al. (2013). In addition, the dynamics in convergence might depend also on the initial distribution of opinion as well as on the inter-connectivity of agents, particularly to their ability to reach out to other agents in the model. It would be interesting to see how the network structure of agents' interconnections inter-plays with the opinion trajectory.

By increasing the sentiment noise in the matching process, we were able to generate a smooth reduction in the opinion variance and an increase in the meeting efficiency. Particularly striking is the reduction in the variance, exhibiting a smooth pattern of a geometric decay. In particular, the presence of sentiments dramatically increases the meeting efficiency, measured as realized opinion changes in an overall number of meetings that took place, while the level of sentiments does not matter as much. In addition to showing the positive effects of sentiments on the opinion convergence, we have shown that the strong presence of positive sentiments facilitates meetings between diverse agents. As the interaction network becomes denser, opinions are propagated through the network more efficiently, strengthening the converging forces.

Along these lines, our simulation treatments support the notion that society might reach consensus also at lower tolerance levels, as long as agents are willing to meet due to, for instance, feeling strong positive sentiments towards each other. The finding

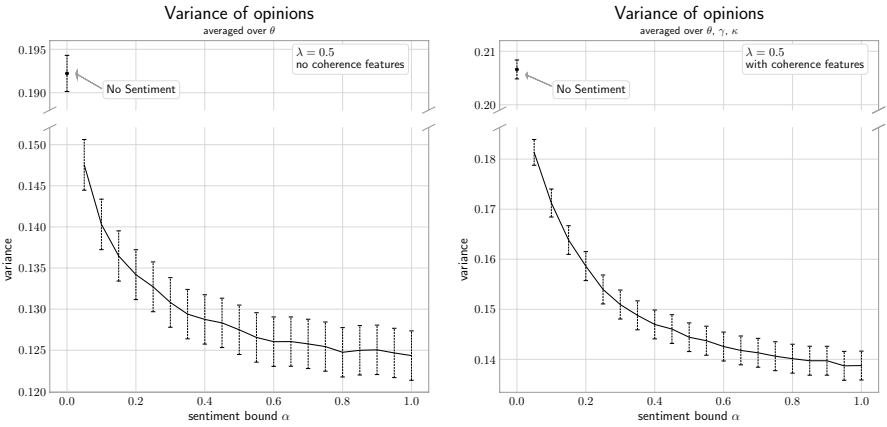

Fig. 9 The opinion variance in $t = 1000$ as a function of the sentiment bound α; left: treatments without coherence features; right: treatments with coherence features

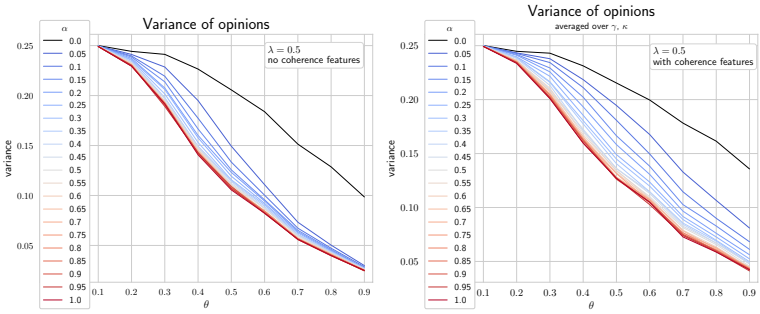

Fig. 10 The variance of opinions by θ, as regulated by sentiment levels. Left: model without coherence features in matching and opinion formation; right: model with coherence features

complements simulation-based study in Steinbacher and Steinbacher (2019), where authors linked a sharp reversal towards extremism at the presence of higher tolerance, suggesting to the existence of the Paradox of Tolerance (Popper, 1945), according to which a highly tolerant society could be seized by the intolerant members.

In addition to the main finding is the discovery of the potentially positive role of keeping agents alert to valid opinions of their counterparts. Namely, at some point, the hyper in-confident agents appeared, who are able to facilitate the appearance of opinion clusters that might push the extreme poles of the opinion spectrum wide-apart and drag the most tolerant agents alongside this route. In particular, when a combination of hyper in-confident agents, high levels of tolerance, and low sensitivity to the validity of opinions is present, small groups of agents that hold extreme opinions can emerge. It is striking that we do not need hard-coded extremists, stubborn agents or contrarians, but the presence of hyper in-confident agents, i.e. agents with very weak attachment to their beliefs, is enough to endogenously produce fringe groups.

We have not shown, however, if and how the introduction of sentiments might explain the rise of extremists in the society, particularly, when they, on one hand, appeal to the sentiments of others and then, on the other hand, behave as stubborn agents who rarely or never change their opinion (Watts & Dodds, 2007). For instance, in Hu and Zhu (2017), the existence of stubborn agents and mass media was able to produce divided societies. Another important field of study might be to show via simulation treatments the rise of racist resentments and anti-minority sentiments in the modern societies (Hooghe & Dassonneville, 2018; Semyonov et al., 2006). Hopefully, the approach taken in this paper offers some potential also for further research along these lines.

To conclude, by splitting the standard opinion model in two parts, we were able to show that even simple models of social interaction can help us shed some light on the stabilizing nature of the complexity of social interaction. A striking increase in communication emerged in the standard opinion model after agents were allowed to be motivated not only by mere rational choice, but also by noisy sentiments (i.e. emotions). It appears that the addition of randomness increases the search space in the agents' collective quest for consensus-finding and alignment. Moreover, we hope that the here presented results might motivate the arrival of additional research that would further expand models of social interaction by human traits, such as trust, integrity, knowledge, memory, truth, compassion. The first necessary step in this direction would be to look at the models of social interaction through the lens of the matching theory as well. We hope that some of these avenues to our main results might show a valid reference for investments of research also within the broader scope of the theory of social interaction.

5.1 Model Documentation

The algorithm is written in C++ (compiled for 64-bit Visual Studio 12). The source code is available in the following GitHub repository: https://github.com/Mitja-ABM-source/OpiFormSentiments. Original simulation data, as well as Python codes to replicate all figures in the paper are available at https://github.com/CKnoppe/OpiFormAnalysis.

Appendix

Roommate Matching Examples

Preference Rankings

The roommate matching algorithm takes as given the preference lists of agents over all their neighbors. This is exemplified in Fig. 11: given the network structure (LHS of the figure), agents create a ranking of their neighbors (RHS). The task of the algorithm is then to take these preference-rankings as input and determine the best possible match for each agent.

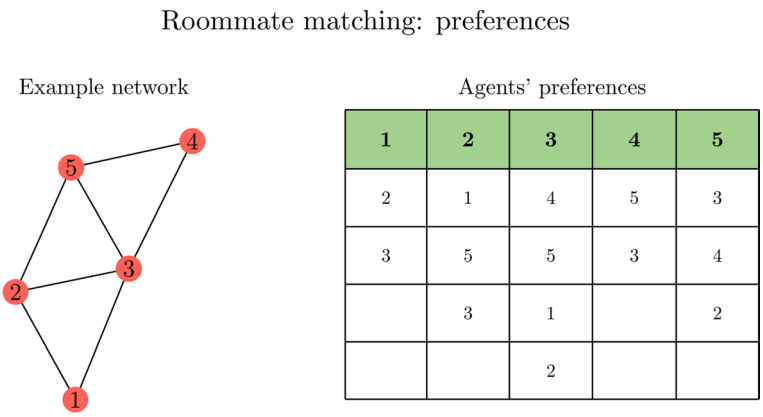

Fig. 11 Agents' preferences (calculated with the RBK, see Sect. 3.2), given a network structure they find themselves in. This is the starting point of the roommate matching algorithm

Phase 1

Given the specific preference ranking, agents are allowed to propose to their favorite choice one after another. This process, which is phase 1 of the algorithm, is depicted in Fig. 12. First, agents 1 and 2 propose to each other, as they are each others favorite neighbors. They form a stable match. For the other three agents, 3, 4 and 5, it is not as simple:

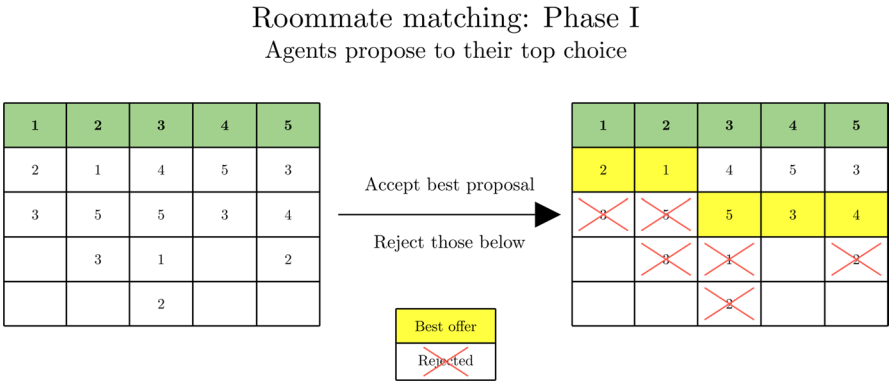

Fig. 12 In phase 1 of the algorithm, agents propose to their favourite neighbors. If they are proposed to, they accept and delete all neighbors from their preference lists that are below the one who proposed. The deletion is mirrored by those that have been deleted. In this example, agents 1 and 2 are each others' favorites and therefore form a stable match by the end of this round. Agents 3, 4 and 5 only hold offers from their second choices, and therefore have reduced lists of length $n > 1$. Only these agents are further processed in the second phase

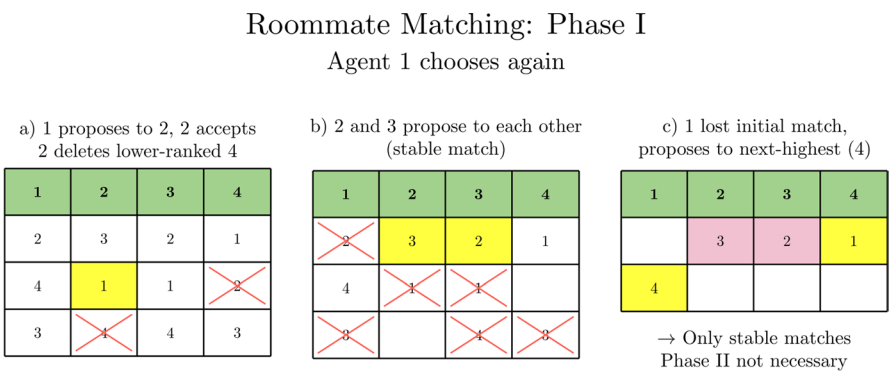

Fig. 13 At first, agent 2 had tentatively accepted agent 1's proposal (**a**). As it then received a better one, the acceptance is revoked however (**b**), so that agent 1 gets to choose again (**c**). In this example it leads to a stable matching in the first round, hence phase 2 of the algorithm would not be necessary

3 proposes to 4, 4 to 5, and 5 to 3. Hence, each holds a proposal from their second-best options and only a tentative acceptance from their first choices. While agents 1 and 2 do not require any further attention, agents 3,4 and 5 therefore move on to phase 2 of the algorithm, which treats cyclical preferences.

Figure 13 shows an example where an initial acceptance is revoked. First, agent 1 proposes to the highest-ranked neighbor, which is agent 2. As agent 2 does not have a better offer yet, she tentatively accepts. As agent 3 proposes to her too however, and agent 3 is ranked higher than 1 in agent 2's preference list, the initial acceptance is revoked. Since agent 1 is now without a potential match, she gets to choose again.

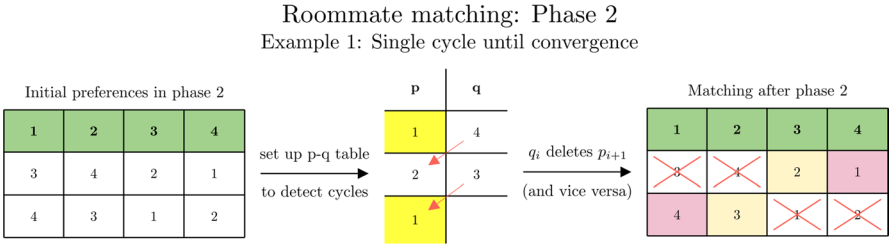

Fig. 14 The second phase starts with agent 1 in position p_0 in this example. The second agent in its reduced preference list is agent 4, which hence becomes q_0. p_1 then has to be the last agent in agent 4's list, i.e. agent 2, q_0 is agent 3, and at p_2, agent 1 appears for the second time. Hence the cycle is completed. Next, agent in the q column delete their least favorite neighbors from their lists. These deletions are, as usual mirrored by the deleted agents in their lists. In this example, it was enough to find one cycle, such that we find stable matches for all agents (1 with 4, and 2 with 3)

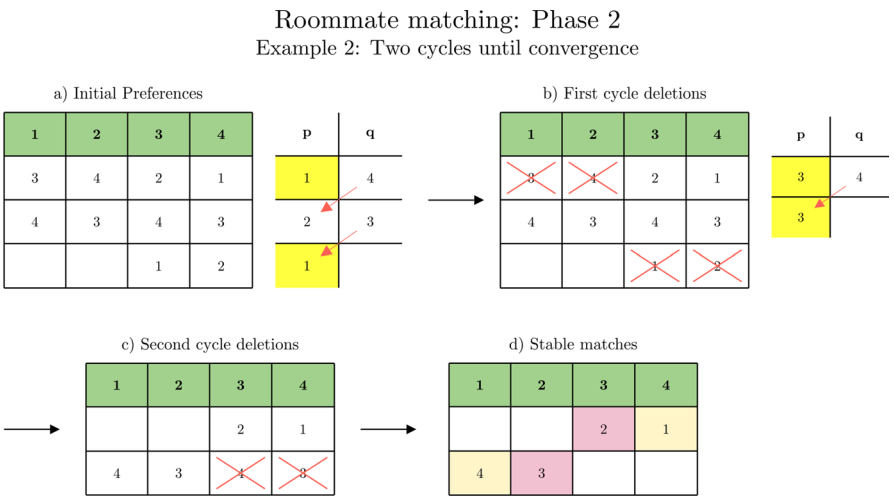

Fig. 15 Example for preferences that require two cycles until stable matches have been found for all agents. In a complete network of these agents, this would be the reduced table after phase 1, if agents 1 and 2 had been each other's least favorite neighbor in the beginning of the round

Phase 2

In the second phase, we set up a p-q table, where p_0 is the root agent, and q_0 is the second-highest agent in p_0's preference list. p_1 is the lowest-ranked agent in q_0's list, and from there we continue. This process is exemplified in Fig. 14: agent 4 is the lowest ranked in agent 1's list. The second pair consists of the lowest-ranked in agent 4's list (agent 2) and the second in its list. I.e. q_i is the second in p_i's list, and p_{i+1} is the last in q_i's list. Once we have detected a cycle, i.e. an agent is found in either of the columns for the second time, all the q-agents delete the lowest-ranked neighbors in their lists. This process typically has to be repeated multiple times, until no agent is left with several neighbors in its preference list anymore (Fig. 15).

Model Algorithm

Algorithm 1 : Treatment Setup

Require: $N \leftarrow 300$ agents AND $t \leftarrow 0, 1, 2, \ldots, 1\,000$ time units;
Ensure: 300 separate C++ objects (i.e. agents) are constructed, such that
expected number of neighbors per agent equals $E(k) \leftarrow \langle k \rangle$;
every agent has at least 1 connection;
every agent i is endowed with specific opinion $x_i : x_i \in \mathbb{R}$;
every agent i is endowed with specific feature vector ϕ_i in Cartesian subspace $\Phi : \Phi \subset \mathbb{R}^4$ and dimensions are represented by independent features $\kappa, \gamma, \theta, \lambda$;
$t \leftarrow 0$;
if $t == 0$ **then**
initialize opinion $x_i(0)$ to each agent i;
end if
if $t < 5$ **then**
initialize independent feature vector $\phi_i (0 \leq t < 5)$ to each agent i;
initialize sentiments ω_{ij}, persistence scores $s_{ij}(0 \leq t < 5)$, such that each neighbor j in the neighborhood $\mathcal{N}(i)$ has a unique preference score $\pi_{ij}(0 \leq t < 5)$ set by agent i;
$t \leftarrow t + 1$;
end if
while $t \leq 1\,000$ **do**
generate time t pairs between agents by using the roommate matching algorithm with preference lists, such that a higher score $\pi_{ij}(t) > \pi_{ik}(t)$ means agent i prefers agent j to agent k (only agents within the same neighborhood can meet);
let each pair (ij) exchange opinion;
obtain opinion differences between agents (ij) after each meeting: $d_{ij}(t) \leftarrow x_j(t) - x_i(t)$;
if $d_{ij}(t) \leq \theta_i \ \wedge \ s_{ij}(t) \geq \gamma_i$ **then**
$x_i(t+1) \leftarrow x_i(t) + \lambda_i(x_j(t) - x_i(t))$;
else$[d_{ij}(t) > \theta_i \ \vee \ s_{ij}(t) < \gamma_i]$
$x_i(t+1) \leftarrow x_i(t)$;
end if
after the meeting, agents obtain new persistence scores $s_{ij}(t)$ of other agents;
after the meeting, agents record new sentiments $\omega_{ij}(t)$ they feel of the other agent from the immediate meeting;
after the meeting, agents assess new preference scores $\pi_{ij}(t)$ of other agents by using new persistence score $s_{ij}(t)$, agreeableness $1 - d_{ij}(t)$, and sentiments $\omega_{ij}(t)$ related to the meeting;
$t \leftarrow t + 1$;
end while

As elucidated by the Algorithm 1, agents will, during the simulation treatments, meet others in their neighborhoods according to the preference matching, whereby multi-featured preference scores $\pi_{ij}(t)$ will determine preference lists, and whereby agents will modify their opinions according to the standard model with bounded confidence from Eq. (3). To facilitate the launch of the matching process, agents are given random scores for the initial five periods of time t.

Acknowledgements Authors would like to thank anonymous referees during the peer-review of the paper, to seminar participants at the Brown Bag Seminar, Kiel University, Germany, May 19, 2022, and to the discussants at the 25th Annual Workshop on Economic Science with Heterogeneous Interacting Agents (WEHIA 2022), Catania, Italy, June 22–24, 2022, for helpful comments and suggestions. All errors remain the responsibility of authors.

Funding Open Access funding enabled and organized by Projekt DEAL. The authors have not disclosed any funding.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Acemoglu, D., Como, G., Fagnani, F., & Ozdaglar, A. (2013). Opinion fluctuations and disagreement in social networks. *Mathematics of Operations Research, 38*(1), 1–27.

Acemoglu, D., & Ozdaglar, A. (2011). Opinion dynamics and learning in social networks. *Dynamic Games and Applications, 1*(1), 3–49.

Acemoglu, D., Ozdaglar, A., & ParandehGheibi, A. (2010). Spread of (mis) information in social networks. *Games and Economic Behavior, 70*(2), 194–227.

Altafini, C. (2013). Consensus problems on networks with antagonistic interactions. *IEEE Transactions on Automatic Control, 58*(4), 935–946.

Arteaga, F., Kapor, A. J., Neilson, C. A., & Zimmerman, S. D. (2022). Smart matching platforms and heterogeneous beliefs in centralized school choice. *The Quarterly Journal of Economics, 137*(3), 1791–1848.

Axelrod, R. (1997). The dissemination of culture a model with local convergence and global polarization. *Journal of Conflict Resolution, 41*(2), 203–226.

Bagnoli, F., Carletti, T., Fanelli, D., Guarino, A., & Guazzini, A. (2007). Dynamical affinity in opinion dynamics modeling. *Physical Review E, 76*(6), 066105.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512.

Battiston, F., Nicosia, V., Latora, V., & Miguel, M. S. (2017). Layered social influence promotes multiculturality in the Axelrod model. *Scientific Reports, 7*(1), 1809.

Blondel, V. D., Hendrickx, J. M., & Tsitsiklis, J. N. (2009). On Krause's multi-agent consensus model with state-dependent connectivity. *IEEE Transactions on Automatic Control, 54*(11), 2586–2597.

Brugnoli, E., Cinelli, M., Quattrociocchi, W., & Scala, A. (2019). Recursive patterns in online echo chambers. *Scientific Reports, 9*(1), 20118.

Büchel, B., Hellmann, T., & Pichler, M. M. (2014). The dynamics of continuous cultural traits in social networks. *Journal of Economic Theory, 154*, 274–309.

Camerer, C. F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Das, A., Gollapudi, S., & Munagala, K. (2014). Modeling opinion dynamics in social networks. *Proceedings of the 7th ACM international conference on web search and data mining* (pp. 403–412).

Deffuant, G., Amblard, F., Weisbuch, G., & Faure, T. (2002). How can extremism prevail? A study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation, 5*(4), 1.

Deffuant, G., Neau, D., Amblard, F., & Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems, 3*, 87–98.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association, 69*(345), 118–121.

Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2017). Modeling confirmation bias and polarization. *Scientific Reports, 7*(1), 40391.

Diakonova, M., Nicosia, V., Latora, V., & San Miguel, M. (2016). Irreducibility of multilayer network dynamics: the case of the voter model. *New Journal of Physics, 18*(2), 023010.

Duggins, P. (2014). A psychologically-motivated model of opinion change with applications to american politics. arXiv preprint arXiv:1406.7770.

Dutta, B., & Sen, A. (2012). Nash implementation with partially honest individuals. *Games and Economic Behavior, 74*(1), 154–169.

Epstein, J.M., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up*. Brookings Institution Press.

Friedkin, N. E., & Johnsen, E. C. (1990). Social influence and opinions. *Journal of Mathematical Sociology, 15*(3–4), 193–206.

Galam, S. (2004). Contrarian deterministic effects on opinion dynamics:"the hung elections scenario''. *Physica A: Statistical Mechanics and its Applications, 333*, 453–460.

Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly, 69*(1), 9–15.

Glaeser, E. L., Sacerdote, B., & Scheinkman, J. A. (1996). Crime and social interactions. *Quarterly Journal of Economics, 111*(2), 507–548.

Glass, C. A., & Glass, D. H. (2021). Social influence of competing groups and leaders in opinion dynamics. *Computational Economics, 58*(3), 799–823.

Golub, B., & Jackson, M. O. (2010). Naïve learning in social networks and the wisdom of crowds. *American Economic Journal, 2*(1), 112–149.

Grimm, V., & Mengel, F. (2020). Experiments on belief formation in networks. *Journal of the European Economic Association, 18*(1), 49–82.

Haas, C., Hall, M., & Vlasnik, S. L. (2018). Finding optimal mentor-mentee matches: A case study in applied two-sided matching. *Heliyon, 4*(6), 1.

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation, 5*(3), 1.

Hegselmann, R., & Krause, U. (2005). Opinion dynamics driven by various ways of averaging. *Computational Economics, 25*(4), 381–405.

Hooghe, M., & Dassonneville, R. (2018). Explaining the trump vote: The effect of racist resentment and anti-immigrant sentiments. *PS: Political Science & Politics, 51*(3), 528–534.

Hu, H., & Zhu, J. J. (2017). Social networks, mass media and public opinions. *Journal of Economic Interaction and Coordination, 12*(2), 393–411.

Irving, R. W. (1985). An efficient algorithm for the "stable roommates'' problem. *Journal of Algorithms, 6*(4), 577–595.

Jadbabaie, A., Lin, J., & Morse, A. S. (2003). Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control, 48*(6), 988–1001.

Kou, G., Zhao, Y., Peng, Y., & Shi, Y. (2012). Multi-level opinion dynamics under bounded confidence. *PLOS ONE, 7*(9), 1–10.

Kozma, B., & Barrat, A. (2008). Consensus formation on adaptive networks. *Physical Review E, 77*(1), 016102.

Kranton, R. E., & Minehart, D. F. (2001). A theory of buyer-seller networks. *American Economic Review, 91*(3), 485–508.

Lejmi-Riahi, H., Belhaj, M., & Ben Said, L. (2019). Studying emotions at work using agent-based modeling and simulation. In *Artificial intelligence applications and innovations: 15th IFIP WG 12.5 international conference, AIAI 2019, Hersonissos, Crete, Greece, May 24–26, 2019, Proceedings 15* (pp. 571–583).

Levy, G., & Razin, R. (2019). Echo chambers and their effects on economic and political outcomes. *Annual Review of Economics, 11*, 303–328.

Li, K., Liang, H., Kou, G., & Dong, Y. (2020). Opinion dynamics model based on the cognitive dissonance: An agent-based simulation. *Information Fusion, 56*, 1–14.
Lorenz, J. (2007). Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C, 18*(12), 1819–1838.
Lorenz, J. (2010). Heterogeneous bounds of confidence: Meet, discuss and find consensus! *Complexity, 15*(4), 43–52.
Macal, C. M., & North, M. J. (2005). Tutorial on agent-based modeling and simulation. In *Proceedings of the Winter Simulation Conference*.
Macy, M.W., & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology* (pp. 143–166).
McCall, J. J. (1970). Economics of information and job search. *The Quarterly Journal of Economics, 84*(1), 113–126.
McVitie, D. G., & Wilson, L. B. (1971). The stable marriage problem. *Communications of the ACM, 14*(7), 486–490.
Noorazar, H. (2020). Recent advances in opinion propagation dynamics: A 2020 survey. *The European Physical Journal Plus, 135*, 1–20.
Pan, Z. (2012). Opinions and networks: How do they effect each other. *Computational Economics, 39*, 157–171.
Patle, A., & Chouhan, D.S. (2013). SVM kernel functions for classification. *2013 International conference on advances in technology and engineering (ICATE)* (pp. 1–9).
Peralta, A.F., Kertész, J., & Iñiguez, G. (2022). Opinion dynamics in social networks: From models to data. arXiv preprint arXiv:2201.01322.
Popper, K. (1945). *The open society and its enemies*. London: Routledge.
Ross, L., & Anderson, C. A. (1982). Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 129–152). Cambridge University Press.
Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of Operations Research, 7*(4), 617–628.
Roth, A. E., & Sotomayor, M. (1992). Two-sided matching. *Handbook of Game Theory with Economic Applications, 1*, 485–541.
Schweitzer, F., Krivachy, T., & Garcia, D. (2019). How emotions drive opinion polarization: An agent-based model. arXiv preprint arXiv:1908.11623.
Semyonov, M., Raijman, R., & Gorodzeisky, A. (2006). The rise of anti-foreigner sentiment in European societies, 1988–2000. *American Sociological Review, 71*(3), 426–449.
Smith, V. L., & Wilson, B. J. (2019). *Humanomics: Moral sentiments and the wealth of nations for the twenty-first century*. Cambridge University Press.
Sobkowicz, P. (2012). Discrete model of opinion changes using knowledge and emotions as control variables. *PLoS ONE, 7*(9), e44489.
Sotomayor, M. (2005). *The roommate problem revisited*. Manuscript, Dept. Econ., Univ. São Paulo.
Steinbacher, M., & Steinbacher, M. (2019). Opinion formation with imperfect agents as an evolutionary process. *Computational Economics, 53*(2), 479–505.
Topa, G., & Zenou, Y. (2015). Neighborhood and network effects. In G. Duranton, V. Henderson, & W. Strange (Eds.), *Handbook of regional and urban economics* (Vol. 5, pp. 561–624). Elsevier.
Urena, R., Kou, G., Dong, Y., Chiclana, F., & Herrera-Viedma, E. (2019). A review on trust propagation and opinion dynamics in social networks and group decision making frameworks. *Information Sciences, 478*, 461–475.
Wang, X., Agatz, N., & Erera, A. (2018). Stable matching for dynamic ridesharing systems. *Transportation Science, 52*(4), 850–867.
Ward, A. J., Sumpter, D. J., Couzin, I. D., Hart, P. J., & Krause, J. (2008). Quorum decision-making facilitates information transfer in fish shoals. *Proceedings of the National Academy of Sciences, 105*(19), 6948–6953.
Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20*(3), 273–281.
Watts, D. J., & Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research, 34*(4), 441–458.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440–442.

Weidlich, W. (1971). The statistical description of polarization phenomena in society. *British Journal of Mathematical and Statistical Psychology, 24*(2), 251–266.
Weidlich, W., & Haag, G. (1983). Opinion formation-an elementary example of semi-quantitative sociology. In *Concepts and Models of a Quantitative Sociology: The Dynamics of Interacting Populations* (pp. 18–53).
Weisbuch, G., Deffuant, G., & Amblard, F. (2005). Persuasion dynamics. *Physica A: Statistical Mechanics and its Applications, 353*, 555–575.
Weisbuch, G., Deffuant, G., Amblard, F., & Nadal, J.-P. (2002). Meet, discuss, and segregate! *Complexity, 7*(3), 55–63.
Wu, Z., Zhou, Q., Dong, Y., Xu, J., Altalhi, A. H., & Herrera, F. (2022). Mixed opinion dynamics based on Degroot model and Hegselmann–Krause model in social networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 53*(1), 296–308.
Zhang, Y., Liu, Q., & Zhang, S. (2017). Opinion formation with time-varying bounded confidence. *PloS One, 12*(3), e0172982.
Zou, W., & Xu, X. (2023). Ingroup bias in a social learning experiment. *Experimental Economics, 26*(1), 27–54.