Market fragmentation and market consolidation: Multiple steady states in systems of adaptive traders choosing where to trade

Aleksandra Alorić[*]
*Scientific Computing Laboratory, Center for the Study of Complex Systems, Institute of Physics Belgrade,
University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia*

Peter Sollich
*Institut fur Theoretische Physik, Georg-August-Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany
and Disordered Systems Group, Department of Mathematics, King's College London, Strand, London WC2R 2LS, United Kingdom*

(Received 19 February 2019; revised manuscript received 8 April 2019; published 21 June 2019)

Technological progress is leading to proliferation and diversification of trading venues, thus increasing the relevance of the long-standing question of market fragmentation versus consolidation. To address this issue quantitatively, we analyze systems of adaptive traders that choose where to trade based on their previous experience. We demonstrate that only based on aggregate parameters about trading venues, such as the demand-to-supply ratio, we can assess whether a population of traders will prefer fragmentation or specialization towards a single venue. We investigate what conditions lead to market fragmentation for populations with a long memory and analyze the stability and other properties of both fragmented and consolidated steady states. Finally, we investigate the dynamics of populations with finite memory; when this memory is long the true long-time steady states are consolidated but fragmented states are strongly metastable, dominating the behavior out to long times.

DOI: 10.1103/PhysRevE.99.062309

I. INTRODUCTION

Whether a consolidated or a fragmented market is more beneficial to a population of traders is a long-standing debate [1–6]. In a consolidated or concentrated market, the majority of trades occurs in one (or a few) as opposed to numerous trading venues. With technological advances we have seen a proliferation of trading venues such as online marketplaces. Even more recently, alternative or dark trading venues have appeared, e.g., dark pools. These are popular not least for their lack of transparency, which makes them interesting for trading large quantities of shares without strongly influencing the price (see, e.g., [6,7]).

The emergence of collective behavior in systems of autonomous agents is a research topic that has seen widespread interest among physicists in the past couple of decades. The main reason for this is the recognition that statistical physics techniques, which contributed to the understanding of macroscopic phenomena arising in large systems of interacting microscopic entities, can be applied to a range of biological, economic, and social systems. A large body of work exists in the physics literature on collective effects in socioeconomic systems [8,9], e.g., mass movement of people [10,11], herd behavior of traders [12], and voting patterns [13,14]. One of the most prominent examples is the minority game, which continues to attract interest due to its simplicity and its ability to reproduce at least "stylized" facts about financial markets [15,16]; extensions of the model also predict interesting grouping phenomena when multiple assets are available to agents [17]. In a similar vein, in this paper, we investigate whether fragmentation and consolidation can arise solely as a consequence of interactions at the level of the agents, combined with individual adaptation.

Some studies of stylized models of market competitions already exist, often pointing out the emergence of monopolies whereby the majority of trades occurs in one trading venue. Pagano [18] argued that when markets are identical (in terms of their transaction costs), risk averse traders will concentrate in a single market. On the other hand, when there is asymmetry, fragmentation might arise with traders being clustered based on the sizes of their desired transactions. Chowdhry and Nanda [2] reached the same conclusion in a system with asymmetrically informed traders and a general number of markets.

Ellison *et al.* [19] and Shi *et al.* [20] also studied competition among markets and the conditions under which such competition can lead either to monopolies or to coexistence of multiple markets. The authors named two significant effects in a competition of double auction trading venues. One of them is the positive size effect, i.e., agents prefer to trade in a market where there are already many traders of the opposite type. As an example, sellers like trading at markets where there are many buyers as this gives them a wider choice of offers. The authors of Refs. [19,20] also suggested the existence of a negative size effect in a double auction market: Agents will prefer being in the minority group of traders more often, with, e.g., buyers benefiting from trading at a market where there are not many other buyers (see, e.g., [21]). Ellison *et al.* [19] pointed out that such negative size effects can enable the coexistence of many markets. On the other hand, Shi *et al.* [20] investigated which of the two effects is stronger and found that

[*] aleksandra.aloric@gmail.com

due to more substantial positive effects, a monopoly will be the favored end state in many situations. The authors of [20] argued that market coexistence remains a possibility when there is strong market differentiation, especially for markets that have different pricing policies: One market might charge a fixed participation fee while another might take a profit fee. A common feature of the studies mentioned above is that some form of Nash equilibrium analysis was used, assuming perfect information about the activity of all traders and maximization of an underlying utility function for the agents.

The increased proportion of trades that take place in dark trading venues—15% of the U.S. share market volume was traded in dark pools in 2013 [22]—suggests increased fragmentation at least between traditional and dark trading. Calling for more research on reasons behind market fragmentation, Gomber *et al.* [23] suggested the heterogeneity of traders and their needs as one of the main drivers of market fragmentation. However, studies of similar effects such as the emergence of market loyalty in fish markets [24], herding [12], and grouping of agents in multiresource minority games [17] show that fragmentationlike phenomena may also be emergent. Nonetheless, existing models for such emergent fragmentation often assume a considerable amount of structure, e.g., in the connectivity among agents, the information available about the actions of other players, the rules of interaction via the market mechanism, the asymmetry between buyers and sellers, etc. In contrast, we study here a model in which initially homogeneous agents adapt only to their private information and show that even in such a system both market fragmentation and consolidation can occur depending on global system parameters.

Based on observations from the CAT tournament [25], where the spontaneous emergence of long-lived market loyalties was seen in complicated systems of adaptive markets and traders, we hypothesize that the reason for fragmentation may not lie in the intricacies of different market mechanisms or trading strategies. Instead, we conjecture that fragmentation is a collective phenomenon arising as a consequence of the continuous adaptation of the individual agents to an evolving system. To test this hypothesis, we developed a stylized model of double auctions and adaptive traders [26,27] that does indeed predict emergent fragmentation under minimal assumptions on the complexity of market and trading mechanisms. The model also shows market consolidation under some circumstances. Our focus in this study is to pin down under what conditions fragmentation and consolidation occur and what relative benefits they bring for the traders. As we will see, the behavior of the model is remarkably rich in spite of its simplicity, with multiple steady states coexisting in the limit of long agent memory. For finite memory length, this can lead to the existence of long-lived metastable states that dominate before the true steady state is reached eventually.

We start with a short description of the model [27] in Sec. II and then proceed to the large memory limit analysis of small systems with $N = 2$ and 4 agents in Sec. III. These can be thought of as two- and four-player games. They are convenient as we can easily track each trader's adaptation. At the same time they already reveal qualitative phenomena related to those we find later in large systems, in particular coordination at the same market (for $N = 2$) and onset of fragmentation via pairwise coordination (for $N = 4$). Moving on to the large population limit ($N \to \infty$), we then first analyze a population with homogeneous buying preferences in Sec. IV A. We develop the relevant mathematical framework and techniques of analysis here and then generalize the results to systems with separate buyer and seller agent types. Finally, we study the system dynamics in some detail to go beyond the steady-state analysis in Sec. V.

Overall we follow a typical statistical physics philosophy in using a model that reduces the underlying market choice dynamics to its key ingredients, allowing us to obtain detailed insights into the origins of the resulting collective behavior. The analysis also relies significantly on statistical physics concepts and methods: We focus mostly on the thermodynamic limit of large agent populations, where we exploit the fact that the behavior of N interacting agents for $N \to \infty$ can be captured by the dynamics of a single agent subject to self-consistently determined population-level order parameters. The main outcome from this physical point of view is the emergence of multiple nontrivial steady states in the large interacting nonequilibrium systems that we study.

II. MODEL

Here we summarize basic assumptions and properties of the model introduced in [26,27], which is the foundation for the analysis in this paper.

Learning. In the model, agents choose among the available markets once in every trading period and submit their order to the chosen market. A key assumption is that agents base their decision of where to trade on their previous experience at the different markets. Agents rely on the following reinforcement rule, which is based on the experience-weighted attraction rule [28,29] but neglects knowledge about the other markets (via so-called fictitious payoffs):

$$A^i_m(n+1) = \begin{cases} (1-r)A^i_m(n) + rS^i_m(n) & \text{for } m \text{ chosen in round } n \\ (1-r)A^i_m(n) & \text{otherwise.} \end{cases} \tag{1}$$

Here $A^i_m(n+1)$ is agent i's attraction to market m at trading period $n+1$ given the agent's score or return $S^i_m(n)$ obtained in the previous trading period (discussed below) and the previous attraction $A^i_m(n)$. To understand the role of r, one can write down the resulting general expression for the attraction at trading round n:

$$A^i_m(n) = \sum_{j=0}^{n-1} r(1-r)^{n-j}\delta_{m^i(j),m}S^i_m(j) + (1-r)^n A^i_m(0),$$

where the Kronecker δ restricts updates to rounds where the agent's chosen market $m^i(j)$ is the one (m) being considered. The factor $r(1-r)^{n-j}$ in this expression is a weight that decays exponentially into the past, becoming small once $n - j$ is of order $1/r$. Thus each agent effectively averages scores over a sliding window into the past of length approximately equal to $1/r$, so $1/r$ can be thought of as setting the length of the agents' memory.

To choose a market at each trading round, an agent translates the learned attractions into probabilities of choosing each markets, using the multinomial logit or softmax function

$$P_m^i(n) = \frac{\exp\left[\beta A_m^i(n)\right]}{\sum_{m'} \exp\left[\beta A_{m'}^i(n)\right]}. \tag{2}$$

This aspect of the model is also in line with the experience-weighted attraction literature [28,29]; β is the intensity of choice and regulates how strongly the agents bias their preferences towards actions with high attractions. For $\beta \to \infty$ the agents choose the option with the highest attraction, while for $\beta \to 0$ they choose randomly and with equal probabilities among all options.

We study agents whose choice of the type of trading order (to buy or to sell) is not adaptive but rather set by a fixed buying preferences p_B^i. This assumption simplifies the analysis while still allowing both consolidation and fragmentation behavior as shown previously [27].

Trading strategies. Agents do not have sophisticated trading strategies in our model and are essentially zero-intelligence traders [30–32]. Their orders to buy (bid) or sell (ask) a single unit of the underlying good at a certain price are independent of previous returns or other information. We assume specifically that bids b and asks a are normally distributed as $a \sim \mathcal{N}(\mu_a, \sigma_a^2)$ and $b \sim \mathcal{N}(\mu_b, \sigma_b^2)$, where we fix $\mu_b > \mu_a$ as in [27]. After each round of trading, each agent receives a score, reflecting their payoff in the trade. This depends on the global trading price set by a chosen market m as well as the order the agent has submitted. The scores of agents who do trade are assigned as in previous studies [30,33]: buyers value paying less than they offered (b) and so their score is $S = b - \pi$. Sellers value trading for more than their ask a and so $S = \pi - a$ is a reasonable model for their payoff; in both cases π is the trading price.

Market mechanism. In the spirit of keeping the model as simple as possible we consider double auction markets in discrete time, counted as before in trading rounds. In every round the global trading price is set by the market: Once all orders have arrived, these are used to determine the average bid $\langle b\rangle$ and average ask $\langle a\rangle$ and set the price

$$\pi = \langle a\rangle + \theta(\langle b\rangle - \langle a\rangle), \tag{3}$$

where θ fixes the price closer to the average bid ($\theta > 0.5$) or the average ask ($\theta < 0.5$), as in [26]. This parameter thus represents the bias of the market towards buyers or sellers. Once the trading price has been set, all bids below this price, and all asks above it, are marked as invalid orders as they cannot be executed at the current trading price. The remaining orders are executed by randomly pairing buyers and sellers. Excess buyers or sellers, i.e., those that cannot be paired, receive zero score, as do the agents who submitted invalid orders.

Note that traders are not informed about the market biases, or the market mechanism in general. The only information they have at their disposal to adapt their market preferences is their personal score.

III. FINITE *N*

A. Two traders: Coordination

To understand collective effects in trading systems, we first build up some intuition by looking at a very simple model with only one buyer and one seller. The traders have a choice between two markets with different biases. As the system consists of only two agents and two markets, fragmentation (or segregation as introduced previously [27]), in which a population will split into distinctive groups favoring one option, is not feasible. However, we can investigate if long-lasting loyalty to a single market emerges, which can signal market consolidation.

To make trading possible the two agents effectively need to coordinate, i.e., to submit orders to the same market. This can lead to one of the agents earning less than they could have done at the other market. One question of interest concerns the conditions under which the agents prefer random decisions of who will be a winner or loser in this manner, as opposed to settling in these roles over longer periods of time. Thus we will focus on the existence of coordination of traders and investigate for which parameter settings agents develop strong preferences for the same market. Intriguingly, this two-player analysis ends up being largely similar to the work by Hanaki *et al.* [34], where a two-agent case was likewise studied as a first step to understanding collective effects. (In [34] these concerned specialization behavior of agents searching for parking spots.)

For the $N = 2$ analysis it is convenient to label the two players as $i = \pm 1$ and similarly for the two markets. We use the following specific parameter settings.

(i) Of the two players, player $i = 1$ always buys while player $i = -1$ always sells ($p_B^1 = 1$ and $p_B^{-1} = 0$).

(ii) Bids and asks are deterministic, i.e., $b \sim \mathcal{N}(\mu_b, 0)$ and $a \sim \mathcal{N}(\mu_a, 0)$, with their difference being fixed to $\mu_b - \mu_a = 1$.

(iii) The trading price at each market is set as defined in [26], $\pi_m = \langle a\rangle + \theta_m(\langle b\rangle - \langle a\rangle)$.

(iv) We assume that the market biases are symmetric, $(\theta_1, \theta_{-1}) = (\theta, 1 - \theta)$, where $\theta \in [0, 0.5]$.

The simplification over our previous work [26,27] of making bids and asks deterministic allows us to focus solely on the coordination of the market choices and does not change the behavior of the system qualitatively. The deterministic order prices then also make the trading prices deterministic: $\pi_m = \mu_a + \theta_m(\mu_b - \mu_a) = \mu_a + \theta_m$.

We can summarize the attraction update rule (1) as

$$A_m^i(n+1) = (1 - r)A_m^i(n) + rS_m^i(n),$$

with the convention that $S_m^i(n) = 0$ if market m was not chosen by agent i in round n. This generalized score is fully determined by the market choice of the opposite player

$$S_m^i(n) = \delta_{m^i(n),m}\delta_{m^{-i}(n),m}\Sigma_m^i, \tag{4}$$

where $m^{(-)i}(n)$ denotes the market of choice of the (co-)player $(-)i$ during trade n and

$$\Sigma_m^i = \begin{cases} \mu_b - \pi_m = 1 - \theta_m, & i = 1 \\ \pi_m - \mu_a = \theta_m, & i = -1 \end{cases} \tag{5}$$

encodes the relevant nonzero score values that depend on the type of market and agent. The logit assignment (2) by which agents choose a market m simplifies for $N = 2$ to

$$P^i_m(n) = \frac{1}{1 + \exp[-\beta m \Delta^i(n)]} = \sigma_\beta(m\Delta^i(n)),$$

where $\sigma_\beta(z) = [1 + \exp(-\beta z)]^{-1}$ is the logistic sigmoid. The choice probabilities do not depend on the attractions to the two markets individually but only on their difference $\Delta^i = A^i_1 - A^i_{-1}$. The latter is updated as

$$\begin{aligned}\Delta^i(n+1) &= A^i_1(n+1) - A^i_{-1}(n+1)\\ &= rS^i_1(n) + (1-r)A^i_1(n)\\ &\quad - \big[rS^i_{-1}(n) + (1-r)A^i_{-1}(n)\big].\end{aligned}$$

The stochastic variable $\Delta^i(n+1)$ thus depends on the choices the agents make in trading round n, $m^i(n)$ and $m^{-i}(n)$, which are drawn from distributions that depend on $\Delta^i(n)$ and $\Delta^{-i}(n)$. This situation simplifies in the long memory limit $r \to 0$, where the attraction differences change sufficiently slowly to average out stochastic fluctuations. One can then effectively replace $\delta_{m^i(n),1}$ by its expected value $\sigma_\beta(\Delta^i(n))$ (and similarly for $-i$ and other market choices) in the score (4). This gives

$$\begin{aligned}\Delta^i(n+1) = r\big[&\sigma_\beta(\Delta^i(n))\sigma_\beta(\Delta^{-i}(n))\Sigma^i_1\\ &- \sigma_\beta(-\Delta^i(n))\sigma_\beta(-\Delta^{-i}(n))\Sigma^i_{-1}\big]\\ &+ (1-r)\Delta^i(n),\end{aligned}$$

which can be further simplified into

$$\begin{aligned}&\frac{\Delta^i(n+1) - \Delta^i(n)}{r}\\ &\quad= -\Delta^i(n) + \big[\sigma_\beta(\Delta^i(n))\sigma_\beta(\Delta^{-i}(n))\Sigma^i_1\\ &\qquad - \sigma_\beta(-\Delta^i(n))\sigma_\beta(-\Delta^{-i}(n))\Sigma^i_{-1}\big].\end{aligned}$$

The finite difference on the left-hand side becomes a derivative in the limit of small r if we switch to the rescaled time $t = nr$, for which a unit time interval corresponds to $1/r$ trading periods:

$$\begin{aligned}\partial_t \Delta^i(t) = -\Delta^i(t) + \big[&\sigma_\beta(\Delta^i(t))\sigma_\beta(\Delta^{-i}(t))\Sigma^i_1\\ &- \sigma_\beta(-\Delta^i(t))\sigma_\beta(-\Delta^{-i}(t))\Sigma^i_{-1}\big].\end{aligned}$$

A convenient change in variables that simplifies this pair of differential equations is $\Delta^1(t) = \xi(t) + \rho(t)$ and $\Delta^{-1}(t) = \xi(t) - \rho(t)$, which after some algebra and exploiting the market symmetry gives

$$\begin{aligned}\partial_t \xi(t) &= -\xi(t) + \frac{1}{2}\frac{\sinh[\beta\xi(t)]}{\cosh[\beta\xi(t)] + \cosh[\beta\rho(t)]},\\ \partial_t \rho(t) &= -\rho(t) + \frac{1-2\theta}{2}\frac{\cosh[\beta\xi(t)]}{\cosh[\beta\xi(t)] + \cosh[\beta\rho(t)]}.\end{aligned} \quad (6)$$

Note that $\xi = (\Delta^1 + \Delta^{-1})/2$ describes the average of the attraction differences of the two agents, while $\rho = (\Delta^1 - \Delta^{-1})/2$ captures the deviation between them.

To understand the dynamics, we first consider its fixed points, which need to satisfy

$$\begin{aligned}\xi^* &= \frac{1}{2}\frac{\sinh(\beta\xi^*)}{\cosh(\beta\xi^*) + \cosh(\beta\rho^*)},\\ \rho^* &= \frac{1-2\theta}{2}\frac{\cosh(\beta\xi^*)}{\cosh(\beta\xi^*) + \cosh(\beta\rho^*)}.\end{aligned} \quad (7)$$

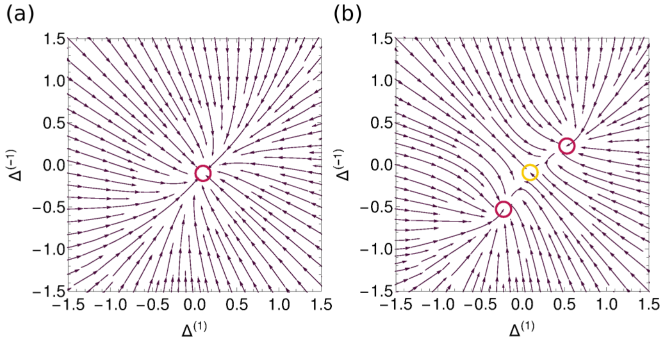

FIG. 1. Two-trader dynamics: flow diagrams (6) for (a) the intensity of choice $\beta = 2$, with a unique fixed point where agents decide largely randomly, and (b) $\beta = 6$, with two new fixed points indicating where coordinated states appear. For the market bias used, $\theta = 0.3$, the critical intensity of choice where coordinated states emerge is $\beta_c = 4.16$.

The first of these equations is always satisfied if $\xi^* = 0$, and in that case the equation for ρ^* has a unique solution whose sign depends on the sign of $1 - 2\theta$. When market 1 is favorable towards buyers ($\theta < 0.5$), ρ^* will be positive. As $\Delta^{\pm 1} = \pm\rho^*$, this can be interpreted as a state where buyers and sellers learn which market is good for them and thus have preferences for opposite markets. (Here Δ^1 is positive, meaning that player 1, the buyer, prefers market 1, which is good for buyers.) As we will see shortly, this solution is only stable for low intensities of choice where the agents' market choice dynamics remains largely random. The intuition for the appearance of an instability with increasing β is that, if agents were to follow through fully on their attractions towards opposite markets, they would never get to trade.

The stability of the solution ($\xi^* = 0, \rho^*$) can be studied by linearizing the dynamical equations (6), resulting in the stability criterion

$$\frac{\beta}{2}\frac{1}{1 + \cosh(\beta\rho^*)} \leqslant 1.$$

Expressed in the original variables Δ^i, the solution with $\Delta^{1*} + \Delta^{-1*} = 0$ is stable as long as

$$\frac{\beta}{2}\frac{1}{1 + \cosh[\beta(\Delta^{1*} - \Delta^{-1*})/2]} \leqslant 1. \quad (8)$$

This stability condition is exactly the same as in Ref. [34] because the learning dynamics we follow is essentially the same and differs only in the details of the deterministic returns.

We illustrate in Fig. 1 that for low intensities of choice, where the stability criterion (8) is satisfied, the fixed point discussed so far is the only one. At higher β the criterion is violated and two new stable fixed points appear. Here the agents' attraction differences are of the same sign, i.e., they prefer going to the same market. This happens even though market 1 favors buyers while market −1 favors sellers.

At first sight it may seem puzzling that for high intensity of choice, one of the agents decides to settle for less in persistently choosing the market where the agent will be awarded lower scores. However, this pattern of behavior in fact maximizes the number of trades that take place. In the

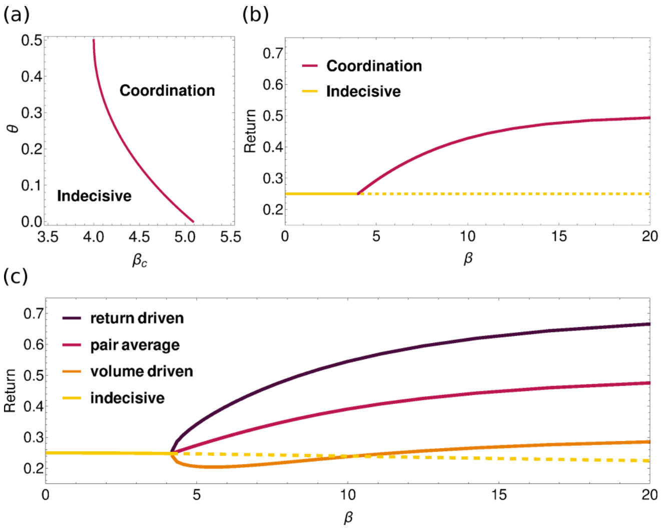

FIG. 2. Two traders: the (θ, β) phase diagram and returns. (a) Coordination and indecisiveness regions for different intensity of choice and market biases (β and θ). (b) Returns for different β in a system with two fair markets $\theta = 0.5$. (c) Returns for different β for market bias $\theta = 0.3$. At the critical $\beta_c = 4.16$, the average return of the two agents in the coordinated state is higher than it would be in the continuation of the low-β fixed point (yellow dashed line), but one of the agents needs to settle for less.

low-β regime, all four pairings of market choices are equally probable, $(m^1, m^{-1}) \in \{(1, 1), (1, -1), (-1, 1), (-1, -1)\}$, but only the first and the last enable trading. On the other hand, in the high-β regime, when both agents persistently choose the same market, they always get to trade, although one of the traders always receives a lower return. For the market parameters used in Fig. 1, $(\theta_1, \theta_{-1}) = (0.3, 0.7)$, the agent who settles for a lower score then receives a score of 0.3, while the other one obtains 0.7. This has to be compared to the average payoff at low β, which by averaging over the four market choice pairings is seen to be $\frac{1}{4}(\theta_1 + \theta_{-1}) = \frac{1}{4}$. Hence both agents clearly earn more in the coordinated regime than by choosing randomly.

We can find the domain of parameters θ and β where the agents will coordinate [Fig. 2(a)] by starting from the regime of agents choosing largely randomly and tracking where the stability condition (8) is first violated as β is increased. As in the case of large populations [27], we observe that β_c increases with increased market difference or bias. The symmetry breaking between markets that coordination requires is therefore not driven by the difference between the markets. In fact, the coordination threshold is lowest for a system with two identical markets ($\theta = 0.5$). One can rationalize this by saying that the agents are happiest to coordinate at one of the markets in this limit as neither needs to settle for less. We show average returns for this setup—a pair of traders choosing between two unbiased markets—as a function of β in Fig. 2(b). One observes the expected average score of 1/4 for low β; as β is increased, the agents effectively realize that coordination at a single market enables more trades and consequently higher average returns.

In Fig. 2(c) we show analogous results for the case of two biased markets ($\theta = 0.3$). We plot the individual agents' payoffs and their average in the state where they coordinate at one market and compare this to the payoff in the largely random low-β state. As a reference we also plot the continuation of the latter to larger β, where it is unstable. It is notable that returns decrease with β on this branch: The more the agents act on their preference for opposite markets, the less often they manage to meet at the same market. This results in more and more trading rounds where both receive a return of zero, dragging down average returns.

By contrast, in the coordinated state the average return increases with β, i.e., as the agents make more and more definite choices. Interestingly, Fig. 2(c) shows that this increase in the average return is accompanied by a growing difference between the returns of the individual agents. These payoff differences can occur in our model because agents are unaware of the opposite player's return, making decisions only on the basis of their own scores. Borrowing terminology from the large system limit [27], we will refer to the agent with the higher return as return driven and the other as volume driven. It is notable in Fig. 2(c) that there is a range of β where the volume-driven agent receives an average return that is lower than not only that of the return-oriented agent, but also the hypothetical return both agents would achieve in the (unstable) uncoordinated state; this regime grows as the markets become more biased.

Intuitively, the return-driven player develops a strong preference for the market where the player can earn more. The other agent will occasionally try the other market, but typically not get to trade there. As this results in a zero return, the player is better off persisting with the coordinated choice, which offers a low but at least nonzero return.

The two-agent systems studied so far can be mapped to two-player games: the symmetric pure coordination game when the markets are unbiased and the battle of the sexes when markets are symmetrically biased. For these games it is known that the two coordinated states correspond to pure Nash equilibria (see, e.g., [35]). In the symmetric pure coordination game, both of these are envy-free (i.e., both agents earn the same), but not so in the battle of the sexes; this is consistent with the differences we saw between unbiased and biased markets, and the Nash equilibria correspond to the $\beta \to \infty$ limit of the coordinated states. There are also mixed Nash equilibria. These correspond to the continuation to $\beta \to \infty$ of our uncoordinated state for the symmetric pure coordination game, but not otherwise. A full correspondence to Nash equilibria could be obtained by modifying the learning rule so that the attractions to markets that were not chosen are kept unchanged. This can be interpreted as fictitious play and is discussed in more detail in [36].

The results described above can be generalized to a pair of traders who do not have strict buyer and seller roles but instead decide to buy with some probability. We assume symmetric preferences for buying, $p_{\mathcal{B}}^1 = 1 - p_{\mathcal{B}}^{-1} = p_{\mathcal{B}}$. For a trade to occur, agents now need to be at the same market and need to submit opposite (buy and sell) orders. As the buying preferences $p_{\mathcal{B}}^i$ are fixed, this only changes Σ_m^i from (5) to

$$\Sigma_m^i = p_{\mathcal{B}}^i\left(1 - p_{\mathcal{B}}^{-i}\right)(1 - \theta_m) + \left(1 - p_{\mathcal{B}}^i\right)p_{\mathcal{B}}^{-i}\theta_m. \tag{9}$$

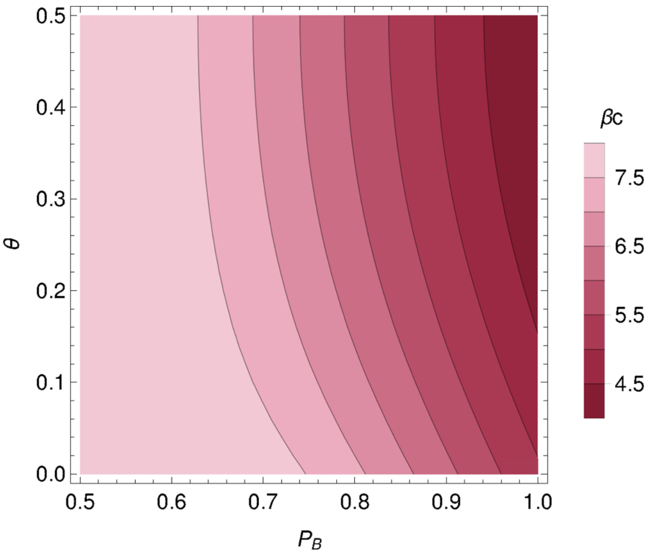

FIG. 3. Two traders: coordination threshold as a function of θ and $p_\mathcal{B}$. Note that the threshold is finite for all system parameters and increases the more similar the agents become, i.e., as $p_\mathcal{B}$ decreases towards 0.5.

To see this, note that agent i receives a buyer payoff $1-\theta_m$ when the agent assumes the role of a buyer (with prob $p^i_\mathcal{B}$) while the opposite player acts as seller; agent i also receives a seller payoff in the opposite configuration. Repeating the calculation above, one then finds that the fixed point conditions (7) both acquire a factor of $p^2_\mathcal{B}+(1-p_\mathcal{B})^2$ on the right-hand side, while the stability condition (8) is multiplied by the same factor. Figure 3 shows contours of the resulting critical β_c for coordination. We note that the coordination threshold increases as $p_\mathcal{B}$ approaches $1/2$: Agents without strong buy and sell preferences need higher intensities of choice to benefit from the coordinated state. This makes sense because agents with $p_\mathcal{B}$ closer to $1/2$ derive a lower benefit from coordinating at a market: As they need to assume buyer and seller roles, trades at the same market happen only with some probability, specifically $p^2_\mathcal{B}+(1-p_\mathcal{B})^2$ in our setting, which approaches $1/2$ for $p_\mathcal{B}\to 1/2$.

B. Four traders: Onset of fragmentation

The two-player system studied above already exhibited an interesting collective phenomenon: coordination at a market to enable more trades, sometimes even to the detriment of an individual agent. Turning to fragmentation, where otherwise homogeneous agents nonetheless learn to adopt different policies, the minimal system size where we can expect a similar effect is $N=4$. We first study two identical buyers and two identical sellers, choosing agents with deterministic buy and sell behavior ($p^i_\mathcal{B}=0$ or 1) for simplicity. A system with four agents is small enough so that we can still easily write down deterministic equations for the evolution of market attractions, but large enough for the first signals of fragmented states to appear as agents can split across the markets in pairs. We consider again symmetrically biased markets, $\theta_1=1-\theta_{-1}=\theta$. As before, the market choice behavior of each agent is determined by their market attraction difference $\Delta^{g,i}$. Here the index g denotes the agent group (buyers or sellers), while i labels agents within each group. For small r the attraction differences again obey deterministic time evolution equations that can be derived by following the reasoning in the preceding section. The only difference lies in the calculation of the return $S^{g,i}_m(n)$ at a chosen market, which now depends on the choices made by all other players:

$$\begin{aligned}S^{g,i}_m(n) = \delta_{m^{g,i}(n),m}\Bigg\{&\frac{\Sigma^g_m}{2}\delta_{m^{g,-i}(n),m}(\delta_{m^{-g,1}(n),m}+\delta_{m^{-g,-1}(n),m})\\ &+\Sigma^g_m(1-\delta_{m^{g,-i}(n),m})(\delta_{m^{-g,1}(n),m}+\delta_{m^{-g,-1}(n),m}\\ &-\delta_{m^{-g,1}(n),m}\delta_{m^{-g,-1}(n),m})\Bigg\},\end{aligned}$$

In this expression, Σ^g_m denotes the deterministic part of the return, which only depends on the chosen market m and the agent type g, by analogy with the two-player case in Eq. (9). The Kronecker δ symbols ensure that other agents are present at the same market m. The first term describes the situation where both agents of the same type go to a single market m; the return is then zero if no agents of the opposite type are at the same market, Σ^g_m if there are two, and $\Sigma^g_m/2$ if there is only one (as our chosen agent then only has probability $1/2$ of being allowed to trade). On the other hand, when the second player of the same type is not at the same market, the player receives the full return if there is at least one trader of the opposite group present. This is described by the second term. The deterministic equations for $r\to 0$ then take the form

$$\partial_t\Delta^{g,i}(t) = -\Delta^{g,i}(t) + \sum_{m=-1}^{1} mS^{g,i}_m(t),$$

where $S^{g,i}_m(t)$ has the meaning of returns averaged over a long time window so that the Kronecker δ's in $S^{g,i}_m(n)$ are replaced with their expected values, exactly as in the derivation for two players. We solve these equations numerically and find that for low and intermediate intensity of choice β the behavior is analogous to that for $N=2$, showing a transition from a single uncoordinated fixed point to two coordinated fixed points as β increases; throughout this range the agents within each group have identical market attractions. The novel feature of the $N=4$ system is that, when β is increased yet further, four new stable states appear. We call these fragmented as each group of agents now "fragments" into two individuals with distinct, and essentially opposite, market preferences. Both markets are then populated by a pair of traders, one from each group. The fragmented fixed points appear in pairs (stable and unstable fixed point) and for high enough value of β unstable fragmented fixed points become partially fragmented, e.g., only one group splits across the markets, while the other group specialize for one market. As these fixed points are not stable we do not show this transition line in Fig. 4.

In Fig. 4 we show the two critical β lines (the coordination and the fragmentation threshold) as a function of the market bias θ, for the above scenario of four players with strict buy and sell roles. The coordination line is very close to the one for two players, which is included for comparison. Both the coordination and fragmentation lines follow the same trend, with the threshold in β increasing as θ departs from 0.5.

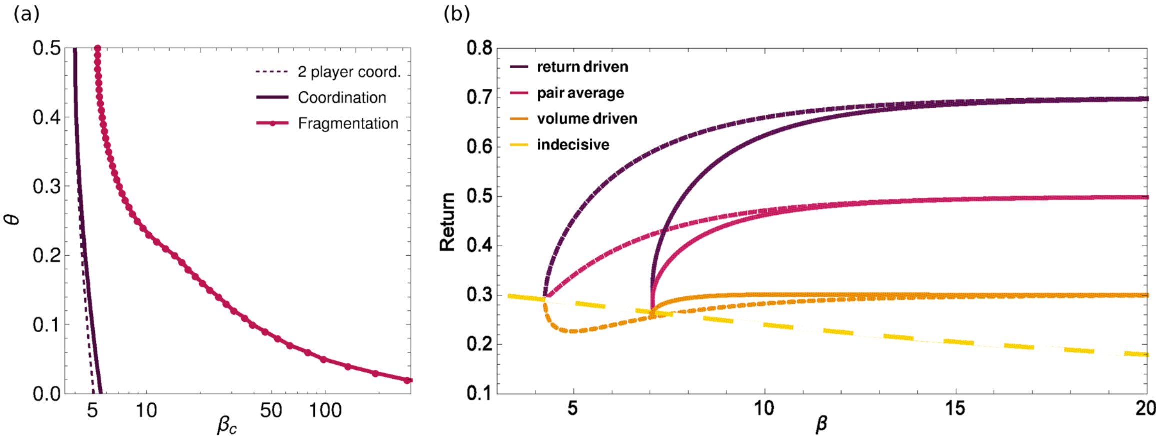

FIG. 4. Four agents (two buyers and two sellers): phase diagram and returns. (a) Phase diagram showing steady states as a function of intensity of choice β and market bias θ. Coordinated steady states exist to the right of the dark violet, solid line and fragmented steady states to the right of the pink solid line with markers. (b) Returns against intensity of choice β for all agents and separately for return- and volume-driven agents; market biases are $(\theta_1, \theta_{-1}) = (0.3, 0.7)$. Dashed lines show coordinated states and solid lines fragmented states. The yellow dashed line shows the average return in the uncoordinated steady state, continued into the instability region at high β.

Figure 4(b) shows return lines for different intensities of choice β: Dashed lines correspond to coordinated states, while solid lines are averages in the fragmented state. Note that the difference in returns in the coordinate state is between groups of agents, with all agents in a group either return driven or volume driven. In the fragmented states there is one return-driven and one volume-driven agent in each group, on the other hand. We note that in the large-β limit the returns achieved in coordinated and fragmented states become identical. This is because with either pattern of market choices, if these choices are made deterministically, then all agents are guaranteed to be able to trade. For finite β, returns in the coordinated state are generally higher than for fragmentation.

Note that the four fragmented states arise because in each agent group there are two ways to assign the two agents to the two markets. For N agents, one therefore expects $\{(N/2)!/[(N/4)!]^2\}^2$ such states. This number grows very rapidly with N, while the number of coordinated states remains at 2.

Finally, as in the analysis of the two-agent system, we can generalize the results by allowing agents to assume the role of buyer with some group-dependent probability $p_{\mathcal{B}}^{(g)}$. We again take these probabilities as symmetric between groups, $p_{\mathcal{B}}^{1} = 1 - p_{\mathcal{B}}^{-1} = p_{\mathcal{B}}$. The deterministic part Σ_m^g of the agents' returns is then modified in a manner directly analogous to Eq. (9), and one can determine the effect on the existence of the various steady states. Figure 5 shows the results for the symmetric markets $(\theta_1, \theta_{-1}) = (0.3, 0.7)$ and symmetric groups. As in the system with only two agents, when the traders' preferences for buying are similar ($p_{\mathcal{B}} \approx 1/2$) they have a weaker incentive to coordinate, resulting in a higher coordination threshold for β (for the sake of clearer visualization we use $1/\beta$ on the y axis). The same behavior is seen also for the fragmentation threshold.

We indicate in Fig. 5 also the regime where a further type of fixed point exists: partially fragmented states. In these states there is a single agent whose market preference is the opposite of that of the other three players, so only one agent group is fragmented. These states evolve for high enough β out of unstable fragmented states, which themselves appear in pairs with the stable fragmented states at the onset of fragmentation. As we will see below, partially fragmented states exist in the large population limit too, though in a limited region of parameter space. In the small system here they are unstable. Intuitively this is likely to be due to the smaller number of trades: In the large-β limit of a partially fragmented fixed point, there will be at most one trade per trading period (only two out of the three agents going to one market will be able to trade), while both fragmented and coordinated states lead to two trades.

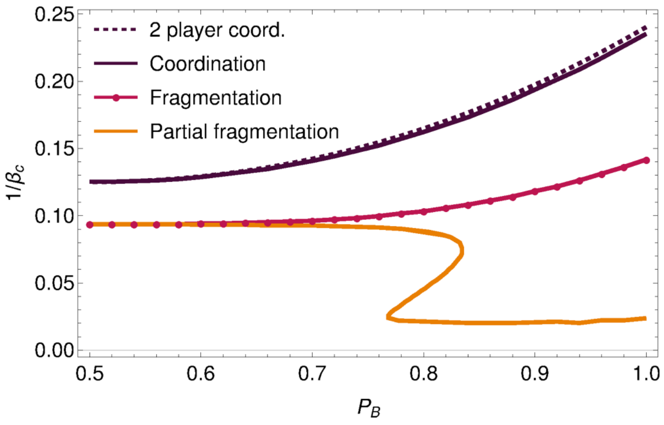

FIG. 5. Four traders: phase diagram when buy and sell roles are probabilistic. Coordination takes place below the dark violet (solid) line and fragmentation below the pink line (solid line with circles). These regions shrink when the difference in the buy and sell preferences of the agents decreases ($p_{\mathcal{B}} \to 0.5$), similarly to the trend in the two-player system (dashed line). Below the orange line partially fragmented fixed points exist, where one of the four agents has a preference for the opposite market.

IV. LARGE POPULATION LIMIT

Population with a fixed buying preference

In studying systems with a small number of agents we have already encountered a rich phenomenology: coordination of agents at a single market, pairwise fragmentation across two markets, and even some mixed states where one group fragments while the other specializes in trading at a single market. We now complement and generalize these results by investigating the possible types of steady state in the large population limit. We start with a simple setting, a population in which all agents have identical preferences for buying $p^i_{\mathcal{B}} = p_{\mathcal{B}} \forall i$. The assumption of population homogeneity is a strong one, but these traders still undergo fragmentation for a broad range of parameters, while the system is simpler to analyze. Thus it is a useful prelude to the analysis of a population consisting of two or more groups with distinct buying preferences.

To describe the steady states of such an initially homogeneous agent population we follow the distribution of attraction differences ($\Delta^i = A^i_1 - A^i_2$) across the population [27]. The state of each market m enters via the probability with which buy ($\mathcal{B}$) and sell ($\mathcal{S}$) orders are executed successfully [27,37]:

$$T_{\mathcal{B}m} = \min\left(1, \frac{Q_{\mathcal{S}m}}{Q_{\mathcal{B}m}D_m}\right),$$

$$T_{\mathcal{S}m} = \min\left(1, \frac{Q_{\mathcal{B}m}D_m}{Q_{\mathcal{S}m}}\right). \tag{10}$$

Here the factors $Q_{\mathcal{B},\mathcal{S}m}$ are the probabilities for submitted buy or sell orders to be valid, i.e., on the right side of the trading price (the explicit expressions are given in Appendix A). Note that, whereas for small systems we simplified to deterministic order prices, we return here to the full model where bids b and asks a are stochastic and the trading price is calculated as in Eq. (3). [As explained in Sec. II, we choose Gaussian distributions for bids and asks, $a \sim \mathcal{N}(\mu_a, \sigma^2_a)$ and $b \sim \mathcal{N}(\mu_b, \sigma^2_b)$; for numerical evaluations we set $\mu_b - \mu_a = 1$ and $\sigma_a = \sigma_b = 1$.] The D_m are demand-to-supply ratios, defined as the number of buyers over the number of sellers at market m. For small r, the attraction difference distribution evolves according to a Fokker-Planck equation

$$\partial_t P(\Delta|p_{\mathcal{B}}, T_\gamma) = -\partial_\Delta[M_1(\Delta|p_{\mathcal{B}}, T_\gamma)P(\Delta|p_{\mathcal{B}}, T_\gamma)] + \frac{r}{2}\partial^2_\Delta[M_2(\Delta|p_{\mathcal{B}}, T_\gamma)P(\Delta|p_{\mathcal{B}}, T_\gamma)], \tag{11}$$

where the drift M_1 and diffusion M_2 both depend on the buying preference of the agents and on the four trading probabilities T_γ. [We use $\gamma = (\tau, m)$ as the generic label for a combination of order type $\tau = \mathcal{B}, \mathcal{S}$ and market choice m.] The drift term is (see Appendix A for details and for the explicit expression of the return distribution from which $\langle S_\gamma \rangle$ is calculated)

$$M_1(\Delta|p_{\mathcal{B}}, T_\gamma) = \sum_{m=-1}^{1} \sum_{\tau\in\{\mathcal{B},\mathcal{S}\}} m p_\tau T_{\tau m} \langle S_{\tau m}\rangle \sigma_\beta(m\Delta) - \Delta, \tag{12}$$

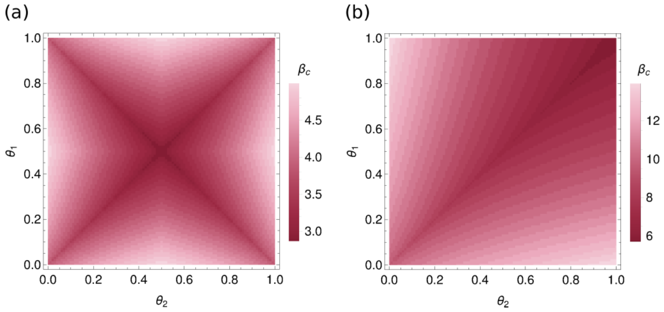

FIG. 6. Critical intensities of choice as a function of market biases for (a) an indecisive population ($p_{\mathcal{B}} = 0.5$) and (b) a population of decisive buyers ($p_{\mathcal{B}} = 0.8$).

where the sum runs over markets m and order types τ and we use the convention $p_{\mathcal{S}} = 1 - p_{\mathcal{B}}$. The strength of the diffusion term is

$$M_2(\Delta|p_{\mathcal{B}}, T_\gamma) = \Delta^2 + \sum_{m=-1}^{1} \sum_{\tau\in\{\mathcal{B},\mathcal{S}\}} \left[p_\tau T_{\tau m}\left(\langle S^2_{\tau m}\rangle - 2m\Delta\langle S_{\tau m}\rangle\right)\right]\sigma_\beta(m\Delta). \tag{13}$$

The steady state of the Fokker-Planck equation is (see, e.g., [38])

$$P(\Delta|p_{\mathcal{B}}, T_\gamma) \propto \frac{1}{M_2(\Delta|p_{\mathcal{B}}, T_\gamma)} \exp\left(-\frac{f(\Delta)}{r}\right), \tag{14}$$

where

$$f(\Delta) = -2\int_0^\Delta d\Delta' \frac{M_1(\Delta'|p_{\mathcal{B}}, D_m)}{M_2(\Delta'|p_{\mathcal{B}}, D_m)} \tag{15}$$

plays a role analogous to a free energy in thermodynamics. When $f(\Delta)$ has a single minimum, $P(\Delta)$ will approach a narrow peak at this location for $r \to 0$ and we have an unfragmented state. Otherwise, as many peaks as there are local minima in $f(\Delta)$ will appear, corresponding to a fragmented state: Each peak represents a subgroup of agents following a distinct market choice strategy.

Note that in the Fokker-Planck description, the market order parameters D_m that determine the trading probabilities T_γ have to be calculated self-consistently from $P(\Delta)$ [27,37]. The same self-consistency condition then also needs to hold at a steady state. Initially we will treat the order parameters as fixed exogenously however. Such a situation could arise if, for example, our agents are just a very small fraction of the overall trading cohort, with the latter fixing the demand-to-supply ratio.

Fragmentation for r → 0

In Fig. 6 we show how the threshold value of the intensity of choice depends on the market biases (θ_1, θ_{-1}) for different agent populations, one indecisive ($p_{\mathcal{B}} = 0.5$) and one made up of decisive buyers ($p_{\mathcal{B}} = 0.8$); for this calculation we set the order parameters to their endogenous value following the self-consistent procedure outlined in (19). We see that for every pair of market biases there is a finite threshold β_c above which fragmentation sets in. When agents are indecisive with regard to buying and selling, the region where fragmentation occurs is greatest when markets are identical or symmetrically biased. For

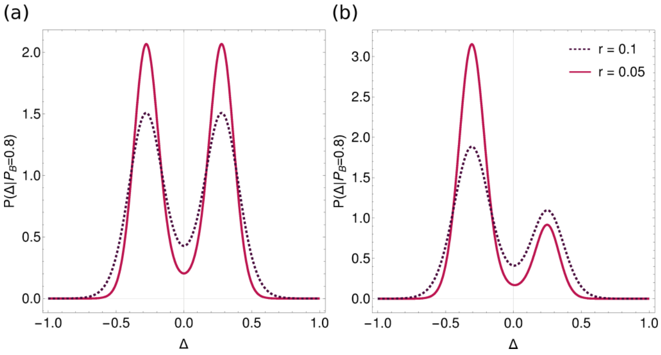

FIG. 7. Steady-state attraction difference distributions of decisive buyers ($p_{\mathcal{B}} = 0.8$). We compare steady states at $\beta = 20$ for (a) two unbiased markets $(\theta_1, \theta_{-1}) = (0.5, 0.5)$ and (b) two symmetrically biased markets $(\theta_1, \theta_{-1}) = (0.3, 0.7)$, for $r = 0.1$ (dashed dark violet line) and 0.05 (solid pink line). The distributions are strongly and weakly fragmented, respectively: on the right, the relative height of the lower peak decreases as r is reduced.

decisive buyers ($p_{\mathcal{B}} = 0.8$), on the other hand, the fragmentation threshold is lowest when the markets are identical. Intermediate values of $p_{\mathcal{B}}$ provide a smooth interpolation between these two situations.

To understand more closely the properties of the fragmented states, we show in Fig. 7 the steady-state distributions of traders with $p_{\mathcal{B}} = 0.8$ when faced with the choice between two unbiased or two symmetrically biased markets. To understand the trend with r, we show the distributions for two different values of r in each case. As expected from (14), the peak width decreases as approximately $\sqrt{r}$ with decreasing r, but in Fig. 7(b) we see that the relative peak heights also depend on r. In fact, if the peaks are located at attraction differences Δ_1 and Δ_2, then the peak height ratio can be written as

$$\frac{P(\Delta_1|p_{\mathcal{B}}, T_\gamma)}{P(\Delta_2|p_{\mathcal{B}}, T_\gamma)} = \frac{M_2(\Delta_2|p_{\mathcal{B}}, T_\gamma)}{M_2(\Delta_1|p_{\mathcal{B}}, T_\gamma)} \exp\left(-\frac{f(\Delta_2) - f(\Delta_1)}{r}\right). \tag{16}$$

This ratio can stay finite for $r \to 0$ only when

$$f(\Delta_1) = f(\Delta_2), \tag{17}$$

and we call this situation strong fragmentation as it survives even in the $r \to 0$ limit. This is the situation in Fig. 7(a). If the free energies at the two peaks are unequal, on the other hand, one continues to have two peaks in $P(\Delta)$ for any nonzero r but the height of one peak decreases (exponentially in $1/r$) as r goes to zero. We call this behavior, which is illustrated in Fig. 7(b), weak fragmentation because the lower peak may become unobservably small for low r; in the strict limit $r \to 0$, the distribution $P(\Delta)$ becomes unimodal again. The strong-weak distinction as defined applies literally only to this $r \to 0$ limit; at nonzero r it becomes a crossover between fragmented states where the emergent subgroups have roughly even (strong) or very different (weak) sizes. At the weakly fragmented state most of the trades happen at a single market (increasingly so as r decreases); we relate this state to market consolidation, and thus the question of fragmentation versus consolidation becomes a question of strong versus weak fragmentation in our setup.

Now that we have a method for finding steady states and classifying them, we return to the space of market order parameters and investigate where fragmentation occurs. In Fig. 8 we show where weakly (colored regions) and strongly (solid lines inside these regions) fragmented states appear, at a fixed intensity of choice $\beta = 8.5$. We compare again indecisive ($p_{\mathcal{B}} = 0.5$) and decisive buyers ($p_{\mathcal{B}} = 0.8$), for three different market setups. We first note that the weak fragmentation region encompasses a very wide range of market conditions (order parameters D_m) for indecisive buyers, but shrinks significantly when the agents have stronger preferences for buying. Looking at the dependence on market setup, an obvious feature is that for two unbiased markets [shown in Fig. 8(c)], equal demand-to-supply ratios ($D_1 = D_{-1}$ line) produce strong fragmentation for both types of agents. This makes sense as the markets are then identical both in their setup $\theta_1 = \theta_{-1}$ and in the prevailing market conditions, making it easy for groups of agents with opposite market preferences to coexist. For the indecisive agents who will act as buyers or sellers with equal probability, the same situation arises when the markets have exactly opposite demand-to-supply ratios ($D_1 = 1/D_{-1}$) and therefore still offer them identical average returns.

With increasing market biases [Figs. 8(a) and 8(b)] the picture obtained for two unbiased markets changes largely smoothly, though note that for decisive buyers (top row) the two crossing lines of strong fragmentation detach into two separate lines [Fig. 8(b), top], with one eventually disappearing out of range.

Market order parameters. So far we have looked at fragmentation behavior driven by exogenously set market conditions (demand-to-supply ratios). We now return to our model as originally set out, where only the adaptive agents we describe trade at the two markets. This leads to the following question: Will a population endogenously create market conditions needed for its fragmentation?

For the case of traders with homogeneous buying preferences $p_{\mathcal{B}}$ this question can be answered relatively straightforwardly. If the steady-state distribution of attraction differences is $P(\Delta|p_{\mathcal{B}}, T_\gamma)$, then the fractions of the whole population buying and selling at market m are

$$N_{\mathcal{B}m} = p_{\mathcal{B}} \int d\Delta\, P(\Delta|p_{\mathcal{B}}, T_\gamma)\sigma_\beta(m\Delta),$$
$$N_{\mathcal{S}m} = (1 - p_{\mathcal{B}}) \int d\Delta\, P(\Delta|p_{\mathcal{B}}, T_\gamma)\sigma_\beta(m\Delta). \tag{18}$$

The demand-to-supply ratio then does not in fact depend on the market preference distribution

$$D_m = \frac{N_{\mathcal{B}m}}{N_{\mathcal{S}m}} = \frac{p_{\mathcal{B}}}{1 - p_{\mathcal{B}}} \tag{19}$$

and is fully determined by $p_{\mathcal{B}}$. In the space of market order parameters in Fig. 8, this endogenous set of market conditions is marked with a black dot. We see that, at high enough β, the population of indecisive buyers fragments strongly when the markets are unbiased or symmetrically biased, and one can check that these results hold independently of the specific market biases used in the figure. Decisive buyers, on the

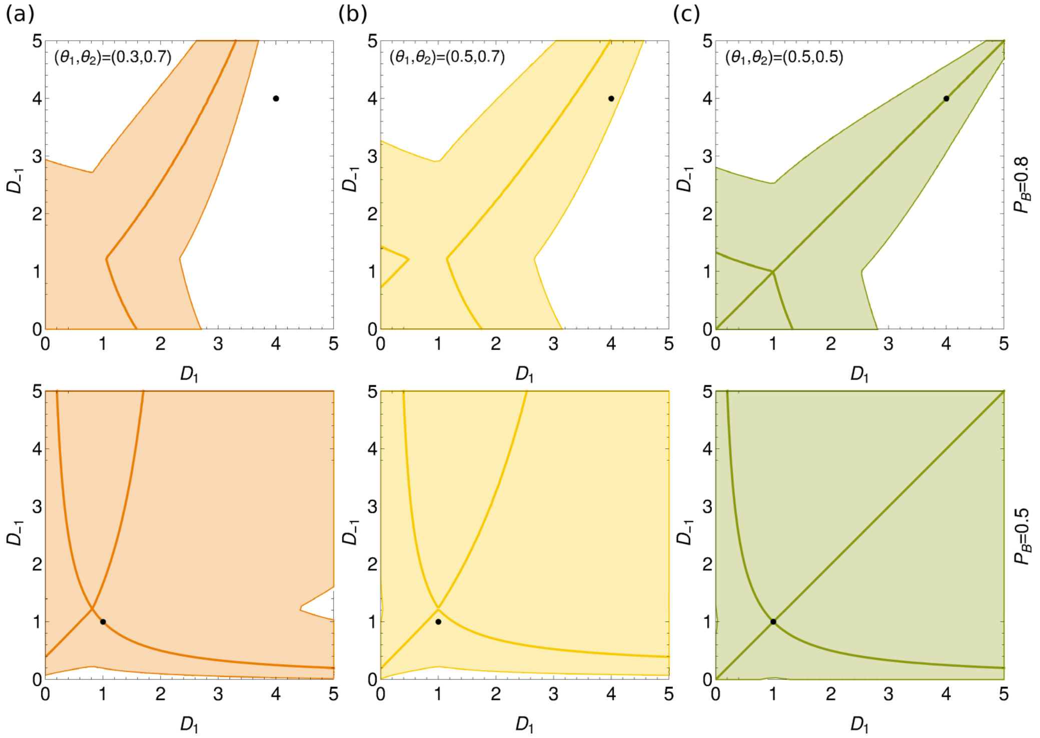

FIG. 8. Single population: steady-state types in the space of market order parameters (D_1, D_{-1}) for $\beta = 8.5$. Shown on top is the population of decisive buyers ($p_B = 0.8$) and on bottom the indecisive population ($p_B = 0.5$) for (a) symmetrically biased markets $(\theta_1, \theta_{-1}) = (0.3, 0.7)$, (b) one unbiased and one biased market $(\theta_1, \theta_{-1}) = (0.5, 0.7)$, and (c) two unbiased markets $(\theta_1, \theta_{-1}) = (0.5, 0.5)$. Colored regions indicate where weakly fragmented states exist (for $r \to 0$). Solid lines inside these regions indicate strongly fragmented states.

other hand, fragment strongly only if the markets are equal (the figure shows only $\theta = 0.5$, but the statement is true for general θ). Otherwise weak fragmentation occurs, although [see Fig. 8(c), top] when the markets are very different at a β above that used in Fig. 8.

With these insights, it is worth revisiting Fig. 6. It shows the existence of the fragmentation threshold β_c for all market biases, and we recall that this threshold is defined as the point where $P(\Delta)$ first acquires two peaks. From what we have seen above, we now understand that for most combinations of market biases, the steady state one finds above β_c is a weakly fragmented one. The exceptions are the dark lines in Fig. 6, which indicate equal ($\theta_1 = \theta_{-1}$) or symmetrically biased ($\theta_1 = 1 - \theta_{-1}$) markets.

V. TWO-GROUP POPULATION

So far in our analysis of the large size limit of a homogeneous population of traders with buying preference p_B, we have shown how for any given pair of market order parameters (D_1, D_{-1}) we can determine the population steady state. We identified three possible types of steady states: unfragmented (U), weakly fragmented (W), and strongly fragmented (S). We now generalize the investigation to populations of agents consisting of groups with different buying preferences. We demonstrate the approach for the case of two groups of the same size, but the principles are general and can be extended to larger numbers of groups or different group sizes. We denote a steady state of a population consisting of two groups by a pair of letters (X,X′). Here X, X′ ∈ {U, W, S} indicates the type of steady state for each group, producing nine different types of population steady states.

We can now find, in the space of market order parameters (D_1, D_{-1}), the domains of different state types as we did in Fig. 8. We can use the figure directly to read off the steady states at $\beta = 8.5$ of a population of two groups with buying preferences $(p_B^{(1)}, p_B^{(2)}) = (0.8, 0.5)$. For example, when the market order parameters are $(D_1, D_{-1}) = (5, 5)$ (the top right corner of all the diagrams), the steady state of the two-group population is (U,W) when the markets are symmetrically biased or biased and unbiased (Fig. 8 left and center) and (S,S) when both markets are unbiased (right diagrams). This simple analysis can be extended to any number of groups because, for market order parameters that are fixed exogenously, the groups are independent.

Our primary interest, however, lies in the case of endogenous market conditions where the agents we model capture the entire trading population and thus define their own market order parameters. In this case, we need to find the steady states self-consistently. We have previously described a procedure for doing this, for nonzero r [27]: Starting from some initial market order parameters (D_1, D_{-1}), one calculates the steady states and updates (D_1, D_{-1}) iteratively, converging eventually to a self-consistent set of order parameters. Here we aim to get a complete picture of all possible steady states, independently of initial conditions. To do this, we start from the update equation for the market order parameters from the iterative approach. These are simply the definitions of

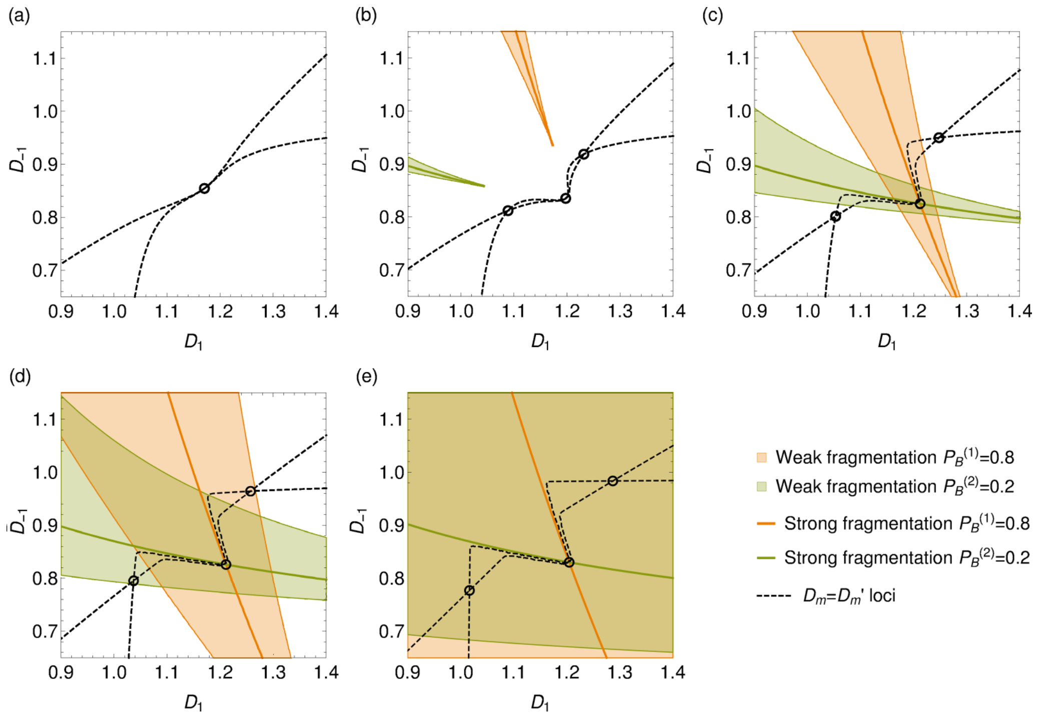

FIG. 9. Steady states of traders with decisive buy and sell preferences: order parameter diagrams. For a two-group system with $(p_{\mathcal{B}}^{(1)}, p_{\mathcal{B}}^{(2)}) = (0.8, 0.2)$, each diagram shows the order parameter self-consistency lines (dashed black) and for each group the weak fragmentation region and the strong fragmentation line; $r = 0.001$ throughout. (a) Single (U,U) solution with $\beta = 1/0.31$, (b) three (U,U) solutions with $\beta = 1/0.29$, (c) one (S,S) and two (U,U) solutions with $\beta = 1/0.265$, (d) (U,W), (S,S), and (W,U) solutions with $\beta = 1/0.245$, and (e) (W,W), (S,S), and (W,W) solutions with $\beta = 1/0.2$. We use the following abbreviations for the steady state of each group: U, unfragmented; W, weakly fragmented; and S, strongly fragmented.

the market order parameters (18) and (19) extended to two groups:

$$D'_m = \frac{N^{(1)}_{\mathcal{B}m} + N^{(2)}_{\mathcal{B}m}}{N^{(1)}_{\mathcal{S}m} + N^{(2)}_{\mathcal{S}m}} = \frac{p^{(1)}_{\mathcal{B}} \int d\Delta\, \sigma_\beta(m\Delta) P\big(\Delta|p^{(1)}_{\mathcal{B}}, T_\gamma\big) + p^{(2)}_{\mathcal{B}} \int d\Delta\, \sigma_\beta(m\Delta) P\big(\Delta|p^{(2)}_{\mathcal{B}}, T_\gamma\big)}{\big(1 - p^{(1)}_{\mathcal{B}}\big) \int d\Delta\, \sigma_\beta(m\Delta) P\big(\Delta|p^{(1)}_{\mathcal{B}}, T_\gamma\big) + \big(1 - p^{(2)}_{\mathcal{B}}\big) \int d\Delta\, \sigma_\beta(m\Delta) P\big(\Delta|p^{(2)}_{\mathcal{B}}, T_\gamma\big)}. \tag{20}$$

We can now define, in the market order parameter space, the two loci where $D'_1 = D_1$ and $D'_{-1} = D_{-1}$, respectively, meaning that one of the order parameters is already self-consistent. The intersection of these loci (two lines, for our case of two markets) then gives us all the self-consistent sets of market order parameters. To distinguish weak and strong fragmentation, the limit $r \to 0$ ought to be taken. To avoid numerical issues we use here instead a small nonzero r to determine the attraction difference distributions $P(\Delta|p_{\mathcal{B}}^{(g)}, T_\gamma)$ from which the D'_m are calculated. In most of what follows we focus on symmetric market setups ($\theta_1 = 1 - \theta_{-1}$) and symmetric agent buying preferences ($p_{\mathcal{B}}^{(1)} = 1 - p_{\mathcal{B}}^{(2)} = p_{\mathcal{B}}$). To avoid having too many parameters to vary, we will fix the market biases to the default values $(\theta_1, \theta_{-1}) = (0.3, 0.7)$ unless stated otherwise.

A. Transitions in populations of decisive and indecisive traders

As shown in previous sections, the intensity of choice β is a crucial parameter determining whether the steady state in a system is fragmented or consolidated. Here we build upon this analysis by investigating how the nature of steady states changes as β is increased. We start this section with examples of steady states of a population with decisive traders $(p_{\mathcal{B}}^{(1)}, p_{\mathcal{B}}^{(2)}) = (0.8, 0.2)$ as well as one with largely indecisive traders $(p_{\mathcal{B}}^{(1)}, p_{\mathcal{B}}^{(2)}) = (0.55, 0.45)$. We then generalize these results to a full phase diagram for the $r \to 0$ limit, giving the number and type of steady states as a function of the intensity of choice β and the buying preference $p_{\mathcal{B}}$.

1. Decisive traders

In Fig. 9 we show, for a series of different β, the market order parameter space (D_1, D_{-1}) with the weak fragmentation region and the strong fragmentation line marked for both groups of a population with $(p_{\mathcal{B}}^{(1)}, p_{\mathcal{B}}^{(2)}) = (0.8, 0.2)$. The order parameter self-consistency lines are also shown.

In Fig. 9(a) we show the low-β regime ($\beta = 1/0.31$), just before the onset of fragmentation. Note that for this β, the steady states of both groups are unfragmented across the entire range of market order parameters shown. The unique intersection of the $D'_m = D_m$ loci identifies a single steady state of type (U,U). Figure 9(b) shows a just slightly increased $\beta = 1/0.29$ where most market order parameters settings still give unfragmented states but there are now three intersections of the self-consistency loci, giving as many (U,U) steady states. In the steady state that is a continuation of the low-β solution the agents show only mild preferences among the markets, with buyers slightly preferring the market that gives higher returns for buyers and similarly for sellers. The other two unfragmented solutions correspond to coordination at one of the two markets so that overall the situation is similar to the one we saw for $N = 2$ and $N = 4$.

Increasing β further [Fig. 9(c)], one crosses the threshold ($\beta_c \approx 1/0.28$ here) where one of the unfragmented solutions first fragments; the continuation of the low-β state is now in the strongly fragmented domain of both groups, while the other two steady states remain unfragmented. Note that since the weak fragmentation regions surround the strong fragmentation lines for both agent groups, there must in fact be a narrow range of slightly-lower-β values where the fragmentation is weak: The low-β solution must change from (U,U) through (W,W) to (S,S) as β increases.

In Figs. 9(d) and 9(e) one observes that with increasing β the fragmentation regions keep growing. This results in the two unfragmented (U,U) solutions changing first into (U,W) and finally (W,W).

We note that inferences about stability from diagrams like Fig. 9 are in general unwarranted; for example, the initial pitchfork bifurcation from one to three (U,U) states in Figs. 9(a) and 9(b) does not necessarily imply that the middle solution is unstable. It would be unstable under repeated updating from D_m to D'_m. However, Eq. (20) shows that this is not the real dynamics but would correspond to a scenario where the dynamics of the order parameters is slowed down artificially so that agents always have time to equilibrate their attraction difference distributions $P(\Delta|p_{\mathcal{B}}^{(g)}, T_\gamma)$ to the current order parameter values.

We highlight one further feature of Fig. 9: For small r as used in the figure, the order parameter self-consistency lines tend to follow segments of the strong segregation lines before emerging on either side into a weak segregation region. This can be understood by noting that the self-consistency line for D_1, for example, is the zero contour of the function $D'_1(D_1, D_{-1}) - D_1$ in the order parameter plane. This function varies steeply as a strong segregation line is crossed, developing discontinuities that look like cliff edges for $r \to 0$. A contour line that hits such a cliff must follow the line of the cliff before returning to the smooth parts of the landscape, which is the effect we see in Fig. 9.

The cliff edges themselves arise because on a segregation line, the free energy function $f(\Delta)$ in Eq. (16) has two minima of equal height. A small change of $O(r)$ in D_1 or D_{-1} will cause similar small changes in the height of these minima, but from Eq. (16) this is enough to cause the weight ratio between the two peaks in $P(\Delta)$ to shift by a factor of order unity. Changes larger than this will transfer all weight from one peak to the other and correspondingly modify D'_1 by a finite amount. For $r \to 0$ the required order parameter changes become infinitesimal, leading to the cliff edge structure of $D'_1 - D_1$ and analogously $D'_{-1} - D_{-1}$.

2. Indecisive traders

We now compare the results above with those for a population consisting of two agent groups with only weak preferences for buying and selling, $(p_{\mathcal{B}}^{(1)}, p_{\mathcal{B}}^{(2)}) = (0.55, 0.45)$. The motivation for this comes from the fact that agents with only mild buy and sell preferences should develop weaker preferences for markets that offer higher returns for buyers or sellers. They will also not be penalized much if only a single group populates a market, as such an arrangement will still sustain a large number of trades.

In Fig. 10 we observe several differences compared to the situation in Fig. 9 for decisive traders, most notably with regard to the number of solutions. Specifically, as can just be discerned from Fig. 10, on crossing the fragmentation threshold four new states appear. These states are also different in nature: They are partially fragmented in the sense that one group of agents is strongly fragmented and thus retains a bimodal distribution of attraction differences for $r \to 0$ while the other is either weakly fragmented or unfragmented. We have seen a similar state in the systems with four agents, although there it was unstable because it reduced the number of possible trades. In the large population limit, having one fragmented and one unfragmented group of agents still leaves many possibilities for trading, especially for indecisive agents where roughly half of each group of agents will probabilistically assume the role of buyer or seller in each trading round. On the general grounds discussed above, the appearance of the partially fragmented (U,S) states is expected to proceed via (U,W) states, though again the β range where the latter appear is numerically small.

When the intensity of choice β is increased beyond that in Fig. 10(a), the low-β solution transitions from (U,U) to (W,W) [Fig. 10(b)] and eventually (S,S) [Fig. 10(d)], i.e., both agent groups fragment first weakly and then strongly. Comparing the partially fragmented solutions in Figs. 10(b) and 10(c), we see that they change from (U,S) to (W,S); finally two of them merge with the uncoordinated (W,W) state into an (S,S) state. The other two partially fragmented states eventually transition into (W,W) states; as in the case of decisive traders, these represent coordination of the agents at a single market.

B. The (β, $p_{\mathcal{B}}$) phase diagram

We have observed both market consolidation and fragmentation when a population is faced with a choice of two symmetric markets, depending on the different choice of system parameters ($p_{\mathcal{B}}, \beta$). We next vary these parameters systematically to construct a detailed phase diagram and study the regions where one finds the various states that we described above. The size of these regions then also gives an indication of how typical the different scenarios are. We continue to focus on symmetric markets with $(\theta_1, \theta_{-1}) = (0.3, 0.7)$ but note that calculations for other (symmetric) market settings give qualitatively similar results. In Fig. 11 we show the phase diagram in the space of intensity of choice β and group preference for buying $p_{\mathcal{B}} \equiv p_{\mathcal{B}}^{(1)}$. This diagram is the large population analog of the diagram for four agents ($N = 4$) shown in Fig. 5. There we had identified regions with states that are unfragmented and indecisive (low β), unfragmented and coordinated, fragmented, and partially fragmented. Broadly, these types of states persist in the large population limit, but they have additional structure that makes for a richer phase diagram.

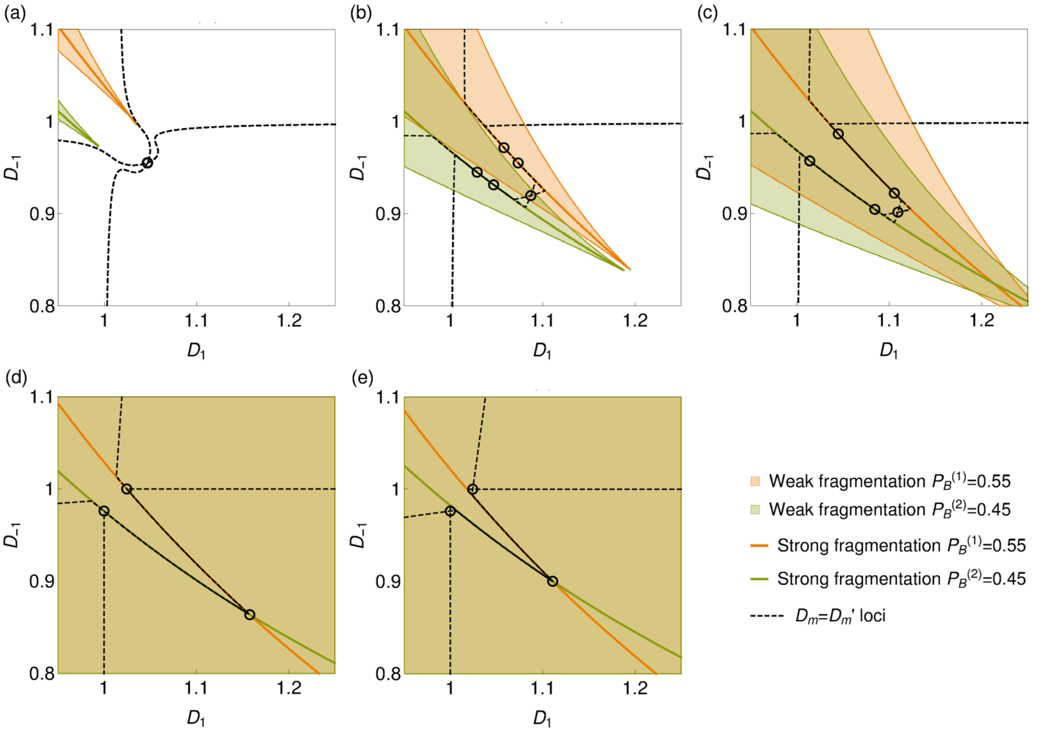

FIG. 10. Steady states of largely indecisive traders: order parameter diagrams. We show the behavior of largely indecisive traders $(p_{\mathcal{B}}^{(1)}, p_{\mathcal{B}}^{(2)}) = (0.55, 0.45)$ for different intensities of choice β in the large memory limit, evaluated numerically for $r = 0.00001$: (a) $\beta = 1/0.31$, one unfragmented solution (U,U); (b) $\beta = 1/0.285$, one weakly fragmented (W,W) and four partially fragmented (U,S) states; (c) $\beta = 1/0.27$ one weakly (W,W) and four partially fragmented (W,S) states; (d) $\beta = 1/0.1$, one strongly fragmented (S,S) and two partially fragmented (W,S) states; and (e) $\beta = 1/0.05$, one strongly fragmented (S,S) and two weakly fragmented states (W,W).

To help visualize the structure of the phase diagram, we show an additional version as an inset that has been distorted to preserve the topology but make even small regions in the phase diagram visible. Also, to avoid having too many separate regions we do not distinguish in the diagram between unfragmented and weakly fragmented states, which both have distributions of market preferences that become unimodal for $r \to 0$. We label such states collectively V to separate them from strongly fragmented states with their bimodal market preference distributions. Two vertical dashed lines mark the two scenarios of decisive and indecisive traders studied above (see Figs. 9 and 10).

We now look in more detail at the structure of Fig. 11. Crossing any line in the phase diagram changes either the number of population solutions or the nature of the steady state for one or both agent groups. We note that, due to the symmetry of the system we consider, many of the changes for the two groups happen simultaneously. In the inset, regions of the parameter space are laid out according to the number of solutions: five solutions on the left and three on the right, with a single solution in the small-β region at the top.

The dark violet line in Fig. 11 is the line where the multiplicity of states changes from 1 to 3 or 5. Looking at the order parameter, self-consistency lines shows that this transition takes place via a pitchfork bifurcation in the former case and two symmetric saddle-node bifurcations in the latter. The dark violet line is an analog of the line shown in the same color in the phase diagram (Fig. 5) of the system with $N = 4$ players. The region of multiple solutions has grown for large N, but the inverse critical intensity of choice $1/\beta_c$ is still an increasing function of $p_{\mathcal{B}}$.

As in Fig. 5, the pink line with circles in Fig. 11 marks the appearance of a steady state where both groups are strongly fragmented. We observe that the critical intensity of choice where this happens diverges ($1/\beta \to 0$) as $p_{\mathcal{B}} \to 0.5$, i.e., the region of strong fragmentation shrinks as the difference between the groups' buying preferences diminishes. Further lines in the phase diagram show where the solution multiplicity changes directly from 3 to 5 (yellow line) and where partial fragmentation occurs (green and orange line) as individual solutions transition from (V,V) to (V,S). Note that in the large population limit such partially fragmented states appear only for populations with moderate preferences for buying, in contrast to the system with $N = 4$ agents (Fig. 5) where they exist for all $p_{\mathcal{B}}$.

We mark one further line (dashed pink) in the main graph of Fig. 11, showing the transition within the small-β (V,V) solution from the unfragmented (U,U) to the weakly fragmented (W,W) state. With this we make an explicit connection to results reported previously (Fig. 7 in [27]) where we investigated the appearance of (weak) fragmentation with increasing intensity of choice. We note that for the system of our first case study $p_{\mathcal{B}} = 0.8$ the thresholds for weak and strong fragmentation almost overlap; the region of the weakly fragmented indecisive state is very narrow for this choice of parameters and in general for $p_{\mathcal{B}}$ above approximately 0.7, while it becomes larger for indecisive traders. The pink circle on the y axis marks the end of the weak fragmentation line.

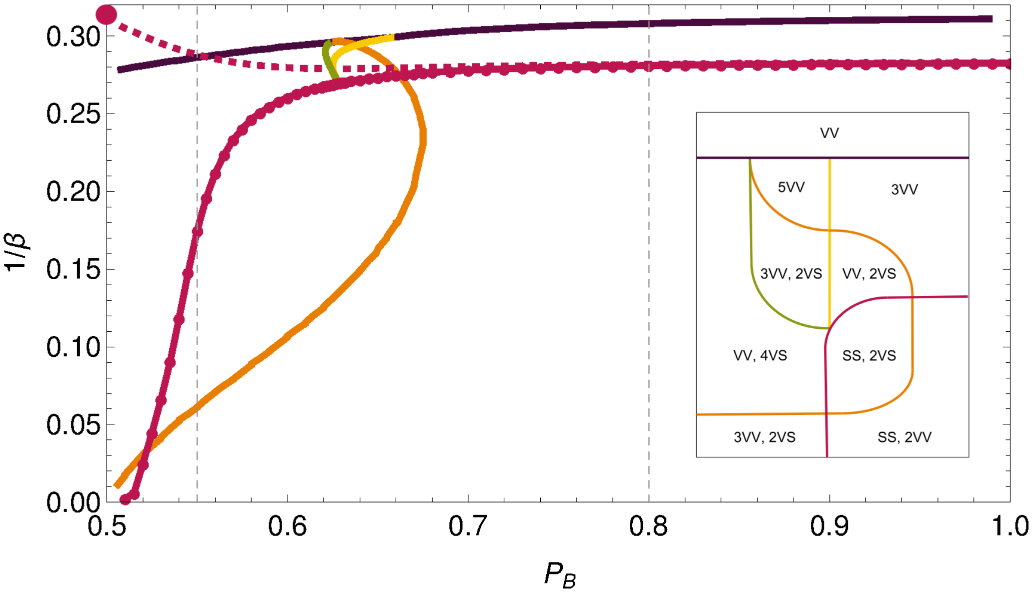

FIG. 11. Types of steady states for two symmetric agent groups, shown in $(\beta, p_{\mathcal{B}})$ space. Crossing each of the lines changes either the number or the type of steady states. The inset shows the distorted but topologically equivalent diagram to show the phase diagram regions more clearly. The dark violet and yellow lines show the change in solution multiplicity, the pink solid line with circles shows the strong fragmentation, the pink dashed line shows the weak fragmentation of the uncoordinated low-β solution, the orange and green lines show the partial fragmentation. Here V denotes a unimodal distribution of agents' market preferences in the $r \to 0$ limit, i.e., an unfragmented (U) or weakly fragmented (W) steady state, and S denotes a strongly fragmented steady state.

It turns out that this is the strong fragmentation threshold of the homogeneous population with even preference for buying and selling ($p_{\mathcal{B}} = 0.5$), with the change from weak to strong fragmentation caused by the additional symmetry between the two groups for this value of $p_{\mathcal{B}}$.

Interestingly, there are two distinct regions in the phase diagram of Fig. 11 where we observe three (V,V) and two (V,S) states, i.e., three unfragmented and two partially fragmented solutions. It turns out that in the region at lower β (higher $1/\beta$) the partially fragmented solutions are coordinated, insofar as both groups of agents have an overall preference for the same market. For high β one has the opposite situation, and it is those uncoordinated (V,S) solutions together with an unfragmented (V,V) solution that then merge into a single (S,S) state as $p_{\mathcal{B}}$ is increased.

We note briefly that the various lines shown in Fig. 11 were detected by solution tracking, e.g., by carefully varying $p_{\mathcal{B}}$ and β and tracking the number and type of solutions; further details can be found in Appendix B. The tracking approach is chosen as it is numerically faster and more reliable than the finite-r procedure we used in previous figures, avoiding, e.g., the numerical noise visible in the two loci in Fig. 10. It is important to remember that the results only provide information about the existence of steady states, not their stability; the latter can be probed only using actual dynamics as discussed below. Figure 11 also relates to fixed market biases so trends with changes in these biases cannot be seen; we have checked, however, that the overall structure of the phase diagram remains intact as long as market biases are symmetric. Quantitative trends were explored in our previous work [27], where we saw that the fragmentation region shrinks as markets become increasingly different.

In summary, the diagram in Fig. 11 shows that for systems with two symmetric markets and two groups of traders with symmetric buying preferences both fragmented and coordinated (or consolidated) steady states exist across a substantial range of values for the intensity of choice β. Single-market dominance happens when the steady state is either unfragmented or weakly or partially fragmented but coordinated: The majority of trades then happens at a single market. On the other hand, markets can coexist, receiving a roughly even share of trades, when the steady state is strongly fragmented or weakly or partially fragmented but uncoordinated. In the former case both markets are visited by both groups, while in the latter case an effective market and group loyalty appears. In the following sections we analyze these different steady states further, with regard to the average population returns they produce and their stability in simulated systems with finite N and r.

C. Average population returns

The phase diagram in Fig. 11 reveals a plethora of possible steady states in the system of two markets and a large population of traders, depending on the traders' learning parameter β and their propensity to act as buyers $p_{\mathcal{B}}$. We now investigate whether these steady states induce differences in average population returns as we saw in small systems, e.g., Figs. 2 and 4. We look at the average population return per trading round, where we count also zero returns that arise from an order being invalid or no trading partner being available.

In Fig. 12 we show average population returns for the two scenarios of decisive $[(p_{\mathcal{B}}^{(1)}, p_{\mathcal{B}}^{(2)}) = (0.8, 0.2)$, Fig. 12(a)] and indecisive $[(p_{\mathcal{B}}^{(1)}, p_{\mathcal{B}}^{(2)}) = (0.55, 0.45)$, Fig. 12(b)] traders. The

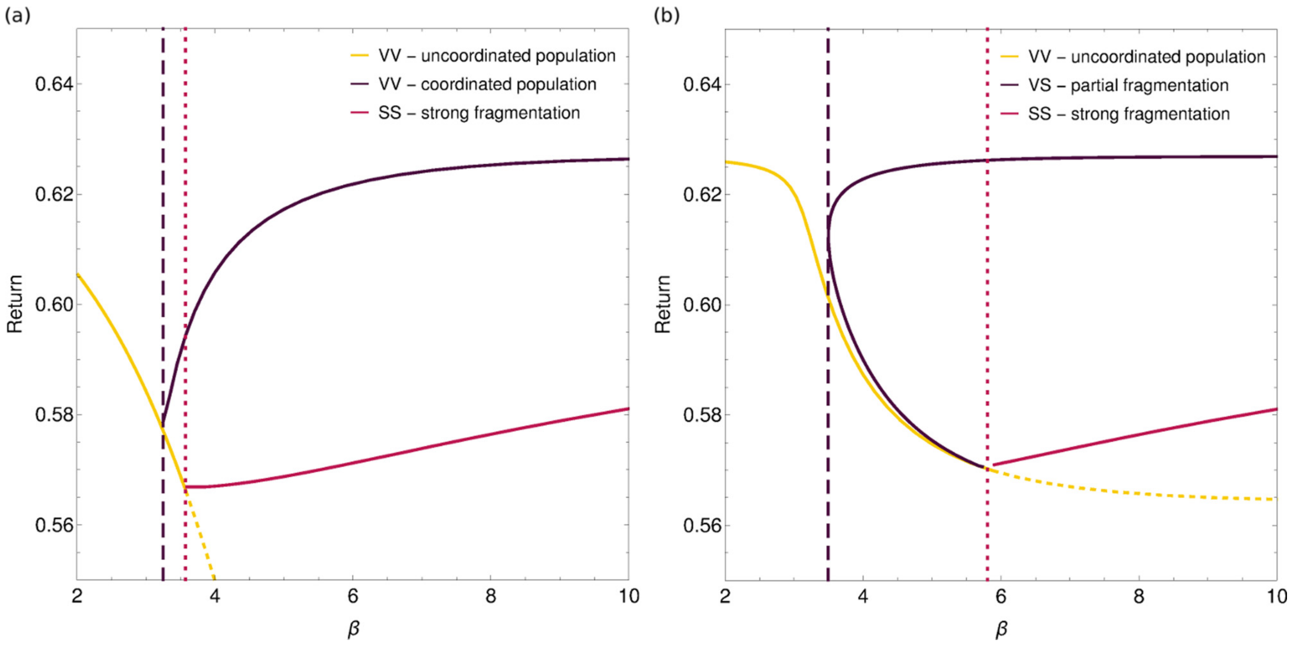

FIG. 12. Average population returns for different steady states in the $r \to 0$ limit. The yellow line shows the low-β steady state representing uncoordinated population (dashed in the regime where it is no longer a bona fide steady state). The dashed dark violet line marks the value of β where the multiplicity of solutions changes (see Fig. 11); at the dashed pink line strongly fragmented steady states first appear. (a) Decisive population $(p_B^{(1)}, p_B^{(2)}) = (0.8, 0.2)$. The dark violet line shows the average population return for a coordinated unfragmented or weakly fragmented steady state and the pink line similarly for a strongly fragmented state. (b) Indecisive population $(p_B^{(1)}, p_B^{(2)}) = (0.55, 0.45)$. The dark violet line gives the average population returns for partially fragmented steady states (coordinated on top, uncoordinated on bottom) and the pink line similarly for a strongly fragmented state.

β dependences reflect the transitions between solution types we saw earlier (in Figs. 9 and 10 and the phase diagram in Fig. 11). The overall trends resemble those for finite N. First, we note that the return of the uncoordinated low-β solution (marked in yellow in Fig. 12) is the lowest among the alternatives once multiple solutions exist. Second, the coordinated states (dark violet) lead to the highest average return. Interestingly, this is not influenced by the type of fragmentation, i.e., it is true for both weakly fragmented and partially fragmented states as long as a majority of the population develops a preference for a single market. By comparison, the strongly fragmented state (pink) always leads to a lower average population return.

The differences in the returns achieved by populations of decisive and indecisive traders, respectively, are driven mainly by the fact that indecisive groups can sustain more trades without requiring the presence of other groups at a market. This is particularly visible in the higher population average return for low β; in this range the decisive population suffers from the group-specific market preferences that tend to separate traders towards different markets and consequently result in a lower number of trades. Additionally, the continuation of the low-β solution is a viable steady state for a broader range of intensities of choice for the indecisive population. The dashed yellow line marks the region of β for which this fixed point is no longer a genuine steady state, as the free energy has multiple minima when evaluated at the order parameters calculated for this fixed point. Along this line the indecisive population return again does not drop as far as it does in the case of a more decisive population.

In Fig. 12(b) we note the occurrence of the saddle node bifurcation in the transition of the indecisive population, with four new (V,S) solutions (which come in two pairs giving identical returns) emerging at once. The top branch corresponds to the average population returns at the coordinated partially fragmented states; for greater values of β (outside the range shown) these states smoothly transition into weakly fragmented, coordinated, states. The bottom branch relates to uncoordinated partially fragmented states that merge into the strongly fragmented (S,S) state for greater β.

Interestingly, the average population return in the high-β limit of the coordinated state also corresponds to the average population return when all traders choose randomly (i.e., $\beta = 0$). This is true because in both limits the average number of agents trading at each market is equal. Intriguingly, this means that when learning is introduced, for low intensities of choice, an agent who makes decisions based on their previous history may be worse off than an agent who plays at random. This effect disappears again only in the large-β limit of the weakly fragmented state, though note that in the latter case one group earns more than the other. Returning to the strongly fragmented state, despite indications that for a given β this is best among the states that do not distinguish between groups in the long run (see Fig. 6 of [27]), in terms of average population return this state is outperformed by random traders ($\beta = 0$).

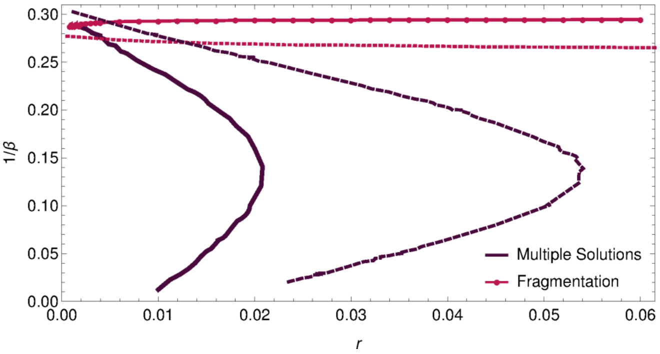

FIG. 13. Memory length dependence of phase boundaries. The pink line (solid with circles and dotted) shows the fragmentation threshold, where at least one steady state is fragmented (weakly or strongly). The dark violet line (dark solid and dashed) shows the boundary of the region where multiple steady states exist. Solid lines represent indecisive traders $(p_B^{(1)}, p_B^{(2)}) = (0.55, 0.45)$ and dashed lines decisive traders $(p_B^{(1)}, p_B^{(2)}) = (0.8, 0.2)$. Multiple steady states exist for small enough r, i.e., long enough memory $1/r$. The market parameters are $(\theta_1, \theta_{-1}) = (0.3, 0.7)$.

D. Dynamics

We now ask what effect the existence of multiple steady states, as predicted by theory for infinite populations, has on the dynamics. We simulate the dynamics numerically, necessarily for finite N and with learning rate $r > 0$, i.e., for finite memory length $1/r$. In previous work we have already shown that the theory predicts the steady-state properties of finite populations quite well (see, e.g., Fig. 4 in [27]). The role of r is more important as this can shift phase boundaries [27]. (Conceptually, the precise distinction for $r \to 0$ between weakly and strongly fragmented states is also lost for $r > 0$ and becomes a crossover.)

In Fig. 13 we illustrate the r dependence of two key phase boundaries for the two populations we have mainly considered so far (decisive $p_B = 0.8$ and indecisive $p_B = 0.55$). We note that the region of multiple steady states shrinks with increasing r for both populations while the fragmentation line is only weakly r dependent. The lines are related to the lines of the same color in the (β, p_B) phase diagram in Fig. 11 and the dashed gray lines marked in Fig. 11 correspond to the $r \to 0$ limit of the (r, β) phase diagram in Fig. 13.

Overall, Fig. 13 tells us that we need to use reasonably small r, certainly below 0.05 for $p_B = 0.8$, to see multiple steady states in numerical simulations. As smaller r slow the dynamics, we choose in practice values of r that are as large as possible while staying well within the multiple states regime.

In Fig. 14 we show numerical data for the actual dynamics of a system of decisive traders $(p_B^{(1)}, p_B^{(2)}) = (0.8, 0.2)$ at our standard market parameters $(\theta_1, \theta_2) = (0.3, 0.7)$, taken from a single run for a population with $N = 2000$ traders (see [37] for simulation details) using the learning rate and inverse decision strength $(r, 1/\beta) = (0.05, 0.16)$. For these parameters the phase diagram of Fig. 13 predicts the existence of three steady states, two weakly fragmented states (with the majority of both groups coordinated at the same market, $m = -1$ or $m = 1$) and a strongly fragmented state (this state was studied in [27]; see Fig. 3 there; it is the unique steady state for the larger $r = 0.1$ used in [27]). As a global summary statistic of the shape of the attraction distributions of the two groups of agents we use the Binder cumulant [39]

$$B = 1 - \frac{\langle \Delta^4 \rangle_{P(\Delta)}}{3\langle \Delta^2 \rangle_{P(\Delta)}^2}$$

and plot this over time (see further discussion in [27,37]). Away from the strongly fragmented state the attraction

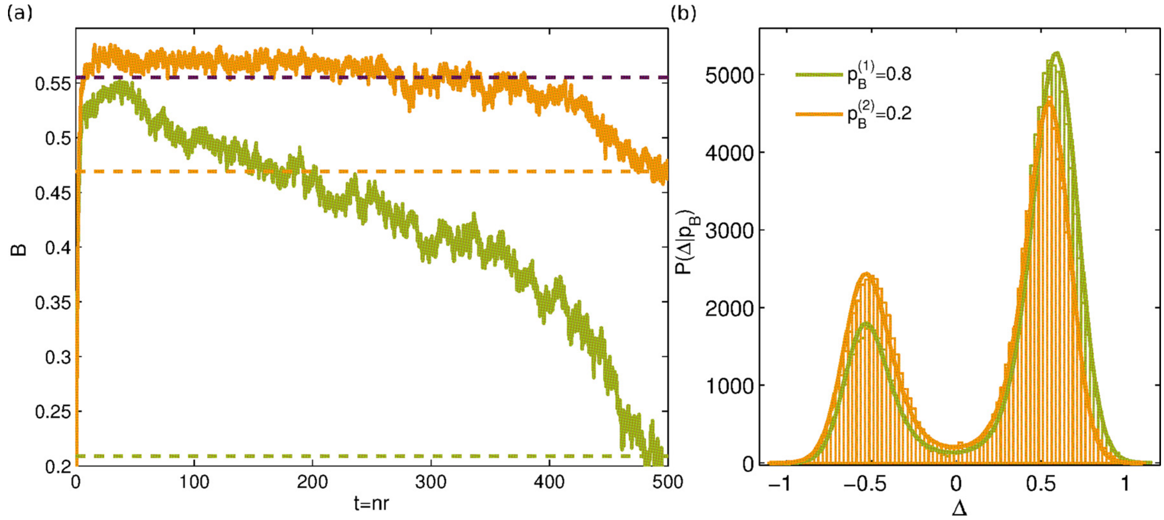

FIG. 14. Metastability of the strongly fragmented state and transition to the weakly fragmented state: dynamical evolution of a system with $N = 2000$ agents using $(r, 1/\beta) = (0.05, 0.16)$, with preferences for buying $(p_B^{(1)}, p_B^{(2)}) = (0.8, 0.2)$, and market parameters $(\theta_1, \theta_2) = (0.3, 0.7)$. (a) Evolution of Binder cumulants of the two attraction distributions of the two agent groups [buyers (green) and sellers (orange)]. Dashed lines are theoretical predictions for the strongly fragmented steady state (dark violet denotes equal for both groups) and weakly fragmented state (green and orange for the two groups). (b) Attraction distributions predicted from theory for the weakly fragmented steady state (solid line) compared to simulation data at $t = 500$ (histogram).

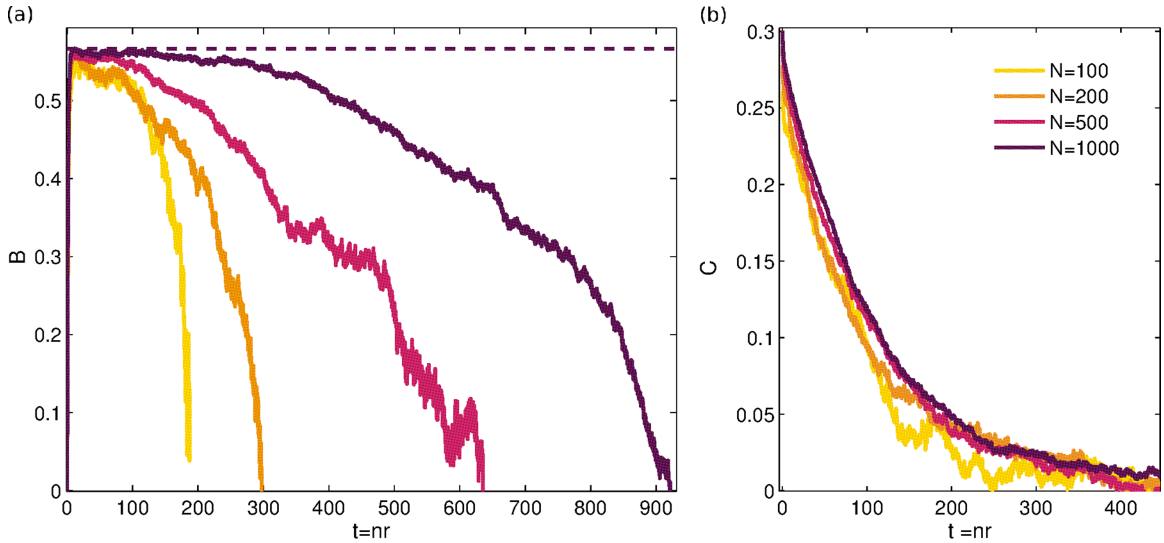

FIG. 15. Lifetime of strongly fragmented state for different system sizes N. (a) Binder cumulant time series (averaged over the two agent groups for compactness) along with the $N \to \infty$ theoretical prediction for the strongly fragmented state (dashed line), showing an increase of lifetime with N. $1/\beta = 0.15$ while the other parameters are as in Fig. 14. (b) Autocorrelation function of single agent attraction differences $C(t) = \langle[\Delta^i(\tau) - \overline{\Delta(\tau)}][\Delta^i(\tau + t) - \overline{\Delta(\tau + t)}]\rangle$. The single agent autocorrelation time is essentially N independent.

distributions of the two groups are not related by a symmetry, so we plot their Binder cumulants separately.

Figure 14 shows that the system quickly reaches the strongly fragmented state, with the Binder cumulants being close to the theoretically predicted value; the slight deviation can be attributed to the finite population size. The dynamics then branches off from the theoretical prediction at $t \approx 50$, showing that the strongly fragmented state is, for finite N, only metastable. The departure is led by one of the agent groups and reaches one of the theoretically expected weakly fragmented states at $t \approx 500$, as shown in Fig. 14 by the agreement of both the relevant Binder cumulants [Fig. 14(a)] and the full attraction distributions [Fig. 14(b)].

We proceed in Fig. 15(a) to analyze the lifetime of the strongly fragmented steady state in more detail. The figure displays Binder cumulant time series for different population sizes at the learning parameters $(r, 1/\beta) = (0.05, 0.15)$ and shows that the lifetime increases with system size (we have not analyzed the N dependence in detail; in the range shown it is approximately linear). We can compare this with the time correlations of the attraction difference Δ for individual agents: Fig. 15(b) graphs this correlation function, measured from the point in time when the strongly fragmented state is first reached. One sees clearly that the single agent correlation time is essentially independent of N, while the lifetime of the strongly fragmented state grows significantly with system size N. The conclusion is that strong fragmentation is a long-lived state of the population for large N, within which single agents effectively "equilibrate" by losing all memory of their initial preferences.

In Fig. 16 we move to the r dependence of the lifetime of the strongly fragmented state, showing Binder cumulants for a small system $N = 200$ for different r values at fixed $1/\beta = 0.15$. For all values of r, rapid initial convergence to the strongly fragmented state is observed. Within this state the Binder cumulants depend weakly on r (as has been noted previously [27]), reflecting the r dependence of the attraction distributions. The lifetime of the strongly fragmented state, set by the decay of the Binder cumulant to lower values, increases with r. This is consistent with the results of Fig. 13, which showed that above some β-dependent threshold value for r the strongly fragmented state is the only steady state and thus must be stable, corresponding to an infinite lifetime. For the value $\beta = 1/0.15$ in Fig. 16, theory predicts this threshold to be $r \approx 0.055$. Numerically, we see that the strongly fragmented state has a finite lifetime up to $r = 0.07$, presumably due to finite population effects for the relatively small $N = 200$ used in the simulations presented in the figure.

We find qualitatively the same features as above also in numerical simulations of the dynamics of a system of indecisive traders, with populations first reaching a long-lived (for large N) strongly fragmented state and eventually decaying

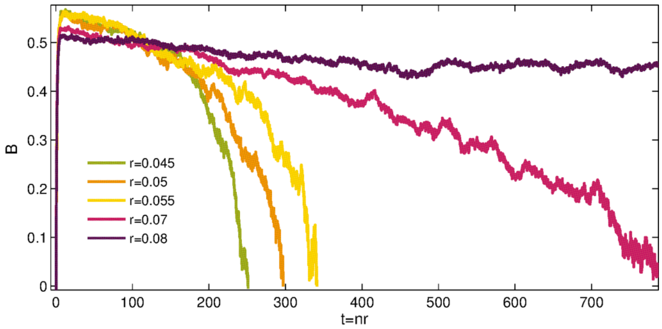

FIG. 16. Binder cumulant time series for different learning rates r at the fixed intensity of choice $1/\beta = 0.15$. All parameters are as in Fig. 14, except for the smaller population size $N = 200$ and r as shown. The lifetime of the strongly fragmented state lifetime increases with r, eventually becoming infinite when this state is the only steady-state solution.

into a partially fragmented state. This is the behavior when such multiple steady states are predicted by our theory, i.e., for small r; for larger r (above $r_c \approx 0.02$; see Fig. 13) strong fragmentation is the only steady state. Quantitatively, we find that where strong fragmentation is metastable its lifetimes are significantly longer than for the decisive traders, exceeding our maximal simulation times of 10^6 trading rounds ($t = 20\,000$) for the largest $r < r_c$.

We comment finally on the role of initial conditions. In the dynamical simulations shown so far we used for these $P(\Delta|p_{\mathcal{B}}) = \delta(\Delta)$, corresponding to the reasonable assumption that the agents have no initial preference for either market. We also explored Gaussian initial distributions for the attraction differences, $P(\Delta|p_{\mathcal{B}}) = \mathcal{N}(\mu, \sigma^2)$. Where there is only a single steady state we then find, as expected, that this state is reached irrespective of the chosen initial condition. On the other hand, where the theory predicts multiple steady states, the initial conditions do matter. We observe that the metastable strongly fragmented state continues to be reached whenever the mean initial attraction difference $|\mu|$ is small enough, irrespective of the standard deviation σ. As $|\mu|$ is increased we see that the dynamics "misses" the metastable strongly fragmented state and rapidly moves to a final weakly or partially fragmented state. This is consistent with the intuition that these states break the symmetry between markets, and hence are favored when the population already starts off with an overall initial preference for one of the markets.

VI. DISCUSSION AND CONCLUSIONS

In this paper, our aim was to investigate the existence of coordinated and fragmented steady states in a system of agents choosing adaptively between two markets. We focused primarily on the long memory limit, where the transition to fragmentation is sharp. We first studied two traders who learn how to coordinate at a market and maximize their average return even though one of them will necessarily earn less. Moving to a four-player system, we observed fragmentation in addition to coordination. Interestingly, we found that coordinated and fragmented states lead to the same average population return for high intensity of choice β, in spite of the presence of two different types of agents (buyers and sellers). In the coordinated state one of the agent types will always earn less, while in the fragmented state both types have the same average, but one agent from each group is less satisfied. Thus, at the fragmented state, average returns do not discriminate between types of agents.

We then introduced a general method for determining the type and number of steady states in the limit of large populations with long memory. This can be done in our setup with only a single order parameter per market. After a preliminary analysis for exogenously determined order parameters, we saw that in the general case a self-consistency criterion determines the order parameters in the steady state. Analyzing a quantity analogous to a free energy then allows one to say whether a population (or one of its groups) is fragmented and whether this fragmentation is strong or weak.

Already for small system sizes we noticed that the agents' preference for buying $p_{\mathcal{B}}$ is an important system parameter. Not only does it influence the critical intensity of choice β on $p_{\mathcal{B}}$ for the onset of fragmentation, but for $N \geqslant 4$ it also qualitatively affects the nature of the steady states. This remains true also for the $N \to \infty$ limit, where we find a rich variety of steady states in the $(\beta, p_{\mathcal{B}})$ diagram, in spite of the simplified nature of our models for markets and traders. These include market coexistence, where both markets attract both types of traders (S,S) and where market–trader specialization occurred (W,W) (uncoordinated weakly fragmented state for moderately indecisive traders); single market dominance (W,W) (coordinated weakly fragmented states); market indifference (U,U) (e.g., for low β); and general vs specialized markets [e.g., (U,S), where a single market attracts both groups of agents while the other can be viewed as specializing towards only one group]. Interestingly, all these different steady states arise without imposing any heterogeneity onto the agents (in contrast to assumptions elsewhere [23]) and fragmentation is the preferred state even when the markets have identical properties (contrary to views expressed in [2,18]).

To interpret our results for the prevalence of fragmentation more broadly we can draw on the work of Cheung and Friedman [40], who used evidence from behavioral game theory to suggest that values of β are consistent across games but increase in more informative environments. The authors also argued that a parameter closely analogous to r increases with the trustworthiness of information in the system. Bearing in mind the results shown in Fig. 13, where for large r and large β the only steady state is the fragmented one, this suggests that more informative environments, or ones where information is more trustworthy because of, e.g., stability over long timescales, might naturally lead to fragmented states. The prevalence of the strongly fragmented state is clear also from Fig. 11, which shows that this state exists for all populations with groups symmetrically biased towards buying and selling, respectively.

One of the nontrivial predictions of our theory is the existence of partially fragmented states, where one group of agents (e.g., those who have a preference for buying) fragments while the other (where agents prefer to sell) does not. We saw that the region in the phase diagram where such states appear increases with N for indecisive traders and shrinks for decisive traders (compare Fig. 5 for $N = 4$ and Fig. 11 for $N \to \infty$).

We studied also the average population returns achieved by agents in the various steady states. For large populations we saw that the coordinated weakly fragmented steady state leads to the highest population average returns, even though one agent group earns less in that state. We also noticed that such steady states, which essentially represent coordination at a single market when $r \to 0$, lead to the same average payoff for large β as for random agents ($\beta = 0$). This is because coordination at a single market, just like random market choice, leads to the same number of buyers and sellers at a single market and thus the same number of successful trades and average returns. Interestingly, this shows that weak learning (finite β) leads to lower returns, e.g. not choosing the strictly best trading venue (in terms of returns) can be worse for an agent than random guessing. This behavior is rather similar to the J-curve effect studied in [41,42] where, in the context of trading agents with different information levels, moderately informed agents earn less from higher informed agents but also from uninformed, randomly trading, agents.

Finally we investigated, by means of numerical simulations, how the theoretically predicted steady states appear in the dynamics of finite agent populations. If the agents start as "blank canvasses" (without initial market preference), we found that the adaptation process always leads to the strongly fragmented state first. This state is metastable, with a lifetime that grows large with population size, and the system eventually settles into one of the weakly fragmented states. This remains true even if there is scatter in the agents' initial preferences, while a systematic initial bias towards one of the markets can cause the dynamics to miss the metastable strongly fragmented state. To put this result into more intuitive terms, two markets that enter into competition to attract on average indifferent traders will always exhibit a period of coexistence in a strongly fragmented state (and if $r > r_c$ this coexistence will last indefinitely), whereas if the population is not indifferent initially then a market monopoly will arise much more quickly.

ACKNOWLEDGMENTS

The authors are grateful to P. McBurney and R. Nicole for useful discussions. P.S. acknowledges the stimulating research environment provided by the EPSRC Centre for Doctoral Training in Cross-Disciplinary Approaches to Non-Equilibrium Systems (Grant No. EP/L015854/1). A.A. acknowledges support from the Ministry of Education, Science, and Technological Development of the Republic of Serbia under Project No. ON171017.

APPENDIX A: DETAILS OF THE FOKKER-PLANCK DESCRIPTION

In this Appendix we provide some of the explicit expressions appearing in the Fokker-Planck description of our market choice model. As per definitions of bid and ask distributions and score assignments, defined in the discussion of trading strategies of Sec. II, the return distributions for an agent choosing a market m and an order type $\mathcal{B}$ or $\mathcal{S}$ are

$$\begin{aligned} P(S|m,\mathcal{B}) &= Q_{\mathcal{B}m}T_{\mathcal{B}m}\frac{1}{Q_{\mathcal{B}m}\sigma_b\sqrt{2\pi}}\exp\left(-\frac{[S-(\mu_b-\pi_m)]^2}{2\sigma_b^2}\right)\theta(S)+\delta(S)(1-Q_{\mathcal{B}m}T_{\mathcal{B}m}),\\ P(S|m,\mathcal{S}) &= \underbrace{Q_{\mathcal{S}m}T_{\mathcal{S}m}}_{\text{agent trades}}\underbrace{\frac{1}{Q_{\mathcal{S}m}\sigma_a\sqrt{2\pi}}\exp\left(-\frac{[S-(\pi_m-\mu_a)]^2}{2\sigma_a^2}\right)\theta(S)}_{\text{non-negative return}}+\delta(S)\underbrace{(1-Q_{\mathcal{S}m}T_{\mathcal{S}m})}_{\text{agent does not trade}}. \end{aligned} \tag{A1}$$

(Note that in statements of these distributions in previous publications [27], μ_a and μ_b were omitted due a typographical error.) When agents have fixed buying preferences $p_\mathcal{B}$, their return distribution is then dependent only on the chosen market m:

$$P(S|m) = p_\mathcal{B}P(S|m,\mathcal{B}) + (1-p_\mathcal{B})P(S|m,\mathcal{S}).$$

The probabilities that an order is valid Q_γ are given by

$$\begin{aligned} Q_{\mathcal{B}m} &= \frac{1}{\sigma_b\sqrt{2\pi}}\int_{\pi_m}^{\infty}db\exp\left(-\frac{(b-\mu_b)^2}{2\sigma_b^2}\right),\\ Q_{\mathcal{S}m} &= \frac{1}{\sigma_a\sqrt{2\pi}}\int_{-\infty}^{\pi_m}da\exp\left(-\frac{(a-\mu_a)^2}{2\sigma_a^2}\right) \end{aligned}$$

and can be expressed in terms of error functions [37].

The transition kernel between two states Δ and Δ' of an agent with buying preference $p_\mathcal{B}$ is

$$\begin{aligned} K(\Delta'|\Delta,p_\mathcal{B}) = \int dS\sum_{m=-1}^{1}&[p_\mathcal{B}P(S|m,\mathcal{B})\\ &+(1-p_\mathcal{B})P(S|m,\mathcal{S})]P(m|\Delta)\\ &\times\delta(\Delta'-mrS-(1-r)\Delta). \end{aligned} \tag{A2}$$

The resulting drift and diffusion terms for small r are discussed in detail in [37]; here (Fig. 17) we provide plots corresponding to Eqs. (12) and (13), evaluated at three different sets of market order parameters for illustration. We consider the value $\beta = 1/0.265$ for the intensity of choice, in order to match Fig. 9(c). The three sets of market order parameters all lie on a horizontal line ($D_{-1} = -1$), while D_1 is changed so that the order parameters lie in the unfragmented, the weakly fragmented, or the strongly fragmented region, respectively. Plots in Fig. 17(a) illustrate market conditions leading to an unfragmented distribution; there is a unique solution of $M_1(\Delta|p_\mathcal{B},T_\gamma)=0$, corresponding to the unique free energy minimum [calculated from Eq. (15) and shown in the bottom row of the figure]. Both are marked by a circle. Figure 17(b) illustrates the weakly fragmented case, where there are three zeros of the drift term (two stable fixed points and one unstable one), corresponding to two minima of the free energy; as the minima are at different heights, the resulting (steady-state) distribution of Δ will become concentrated around the lowest minimum for $r \to 0$, as discussed in the main text. Finally, the case shown in Fig. 17(c) has two equal minima of the free energy and thus represents a strongly fragmented scenario. Note that the diffusion term $M_2(\Delta)$ is in all three cases of order unity and does not affect the number of free energy minima; it only makes a quantitative contribution to the free energy and hence to $P(\Delta|p_\mathcal{B},T_\gamma)$.

APPENDIX B: ALGORITHMIC REMARKS

The method of finding all steady-state solutions by identifying loci of self-consistent market order parameters is the best way to exhaust market order parameter space and thus to find all the solutions for the finite r. By identifying the domains where these solutions lie, we can also fully characterize the solution at nonzero r, obtaining information

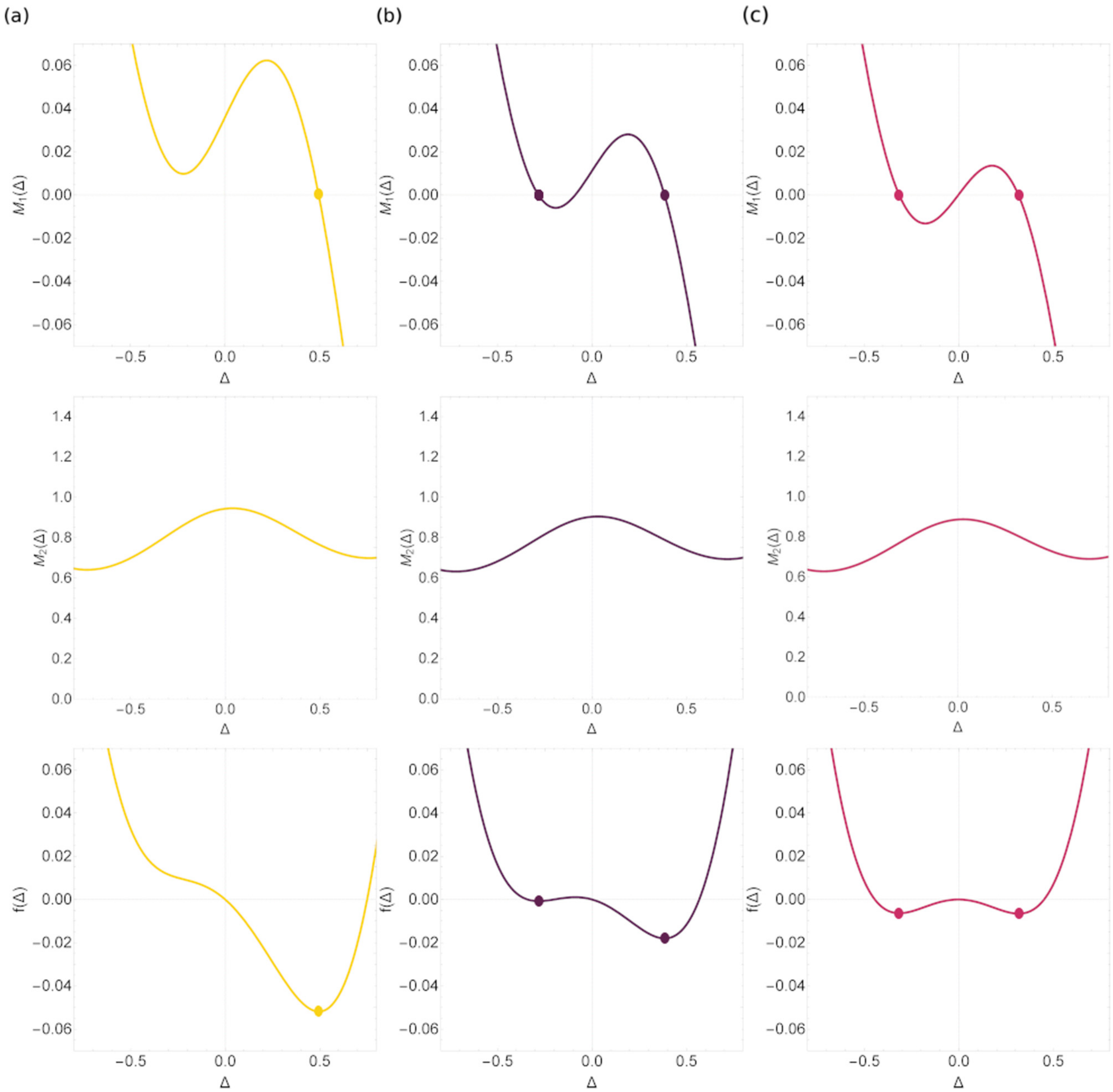

FIG. 17. Drift $M_1(\Delta)$ (top), diffusion $M_2(\Delta)$ (middle), and free energy $f(\Delta)$ (bottom) functions for a subgroup with preference for buying $p_{\mathcal{B}} = 0.8$. Function plots illustrate three qualitatively different conditions for the following pairs of market order parameters: (a) $(D_1, D_{-1}) = (1, 1)$, unfragmented region; (b) $(D_1, D_{-1}) = (1.1, 1)$, weakly fragmented region; and (c) $(D_1, D_{-1}) = (1.15, 1)$, strongly fragmented region. Market biases are set to the standard values $(\theta_1, \theta_{-1}) = (0.7, 0.3)$. All functions are evaluated at the intensity of choice $\beta = 1/0.265$.

about the limit $r \to 0$ by extrapolation. However, this method is numerically demanding as for every point in order parameter space we need to find a steady-state distribution (its normalization usually takes most of the processing time) and recalculate the corresponding order parameters. Checking what corrections arise for $r \to 0$ takes additional time. We describe numerically less demanding alternatives below.

Population with homogeneous market preferences. We have seen that, depending on system parameters, the attraction distribution for a group of agents can be unimodal in the $r \to 0$ limit (U and W states). These states represent a population where the market preferences within the group are homogeneous. This realization offers a straightforward way to find all the states of this type for any system parameter. The demand-to-supply order parameters simplify to

$$D_m = \frac{p_{\mathcal{B}}^{(1)} \int d\Delta\, \sigma_\beta(m\Delta) P(\Delta|p_{\mathcal{B}}^{(1)}) + p_{\mathcal{B}}^{(2)} \int d\Delta\, \sigma_\beta(m\Delta) P(\Delta|p_{\mathcal{B}}^{(2)})}{(1-p_{\mathcal{B}}^{(1)}) \int d\Delta\, \sigma_\beta(m\Delta) P(\Delta|p_{\mathcal{B}}^{(1)}) + (1-p_{\mathcal{B}}^{(2)}) \int d\Delta\, \sigma_\beta(m\Delta) P(\Delta|p_{\mathcal{B}}^{(2)})} = \frac{p_{\mathcal{B}}^{(1)} \sigma_\beta(m\Delta^{(1)}) + p_{\mathcal{B}}^{(2)} \sigma_\beta(m\Delta^{(2)})}{(1-p_{\mathcal{B}}^{(1)}) \sigma_\beta(m\Delta^{(1)}) + (1-p_{\mathcal{B}}^{(2)}) \sigma_\beta(m\Delta^{(2)})}, \tag{B1}$$

where in the right hand side we have used $\langle\sigma_\beta(\Delta)\rangle = \sigma_\beta(\langle\Delta\rangle)$, a relation that is exact in the $r \to 0$ limit where the steady-state distribution is a δ distribution centered at $\Delta^{(g)}$. To identify these peak positions we find the zeros of the first jump moment M_1 as defined in Eq. (12), taking into account the dependence of D_m on the attraction difference $\Delta^{(g)}$ in each group. This means that when searching for a steady state in which both groups of traders have homogeneous market preferences, we need to solve the peak position equations for the two groups simultaneously:

$$\begin{aligned} M_1^{(1)}\big(\Delta^{(1)}\big|p_{\mathcal{B}}^{(1)}, D_m(\Delta^{(1)},\Delta^{(2)})\big) &= 0,\\ M_1^{(2)}\big(\Delta^{(2)}\big|p_{\mathcal{B}}^{(2)}, D_m(\Delta^{(1)},\Delta^{(2)})\big) &= 0. \end{aligned} \tag{B2}$$

Every solution $(\Delta^{(1)*}, \Delta^{(2)*})$ found in this way needs to be checked for consistency with the initial assumption of homogeneous market preferences, i.e., the market order parameters corresponding to every solution pair need to belong to the unfragmented or weakly fragmented solution domain. This is done by calculating the corresponding order parameters D_m from Eq. (B1) and finding the free energy corresponding to these order parameters. If the global free energy minimum is centered at $\Delta^{(g)*}$ the solution is consistent with our initial assumption and we have found a (homogeneous) population steady state. Depending on the signs of Δ^*, we classify such steady states further as either coordinated for $\Delta^{(1)*}\Delta^{(2)*} > 0$ or uncoordinated for $\Delta^{(1)*}\Delta^{(2)*} < 0$. For any finite intensity of choice β, a single agent can of course choose another market even if the state is categorized as coordinated at market 1, but the categorization is exact for the $\beta \to \infty$ limit.

In the second case study ($p_{\mathcal{B}} = 0.55$), the continuation of the low-β fixed point is a solution we can consistently find by this method for a wide range of intensities of choice, much wider than when the groups have more pronounced buy and sell preferences. Crossing the dark violet line in the phase diagram (Fig. 11), the new fixed points that arise all turn out to be inconsistent with the homogeneous population assumption until very high intensities of choice. This is why we need to employ different techniques to find the other solutions presented in Fig. 10. Only when the intensity of choice is increased further do partially fragmented states cease to exist, and solutions consistent with the homogeneous population assumption return.

Strongly cofragmented state (S,S). To find if these states exist we apply a procedure based on a Maxwell construction argument outlined in Sec. IV A and in [37] for a population consisting of a single group. For each group we define a locus in the space of order parameters (D_1, D_2) for which the strong fragmentation condition (8) is satisfied. If there is an intersection (D_1^*, D_2^*) between the two loci there are market demand-to-supply ratios in which both groups favor a strongly fragmented state. We finally need to confirm that the two order parameters can be created if only the two fragmented groups trade on the markets. If we assume the strongly fragmented distributions are of the form

$$P\big(\Delta\big|p_{\mathcal{B}}^{(g)}\big) = \omega^{(g)}\delta\big(\Delta - \Delta_1^{(g)}\big) + (1-\omega^{(g)})\delta\big(\Delta - \Delta_2^{(g)}\big),$$

then the corresponding order parameters are

$$D_m = \frac{N_{\mathcal{B}m}}{N_{\mathcal{S}m}},$$

$$D_m = \frac{p_{\mathcal{B}}^{(1)}\big[\omega^{(1)}\sigma_\beta\big(m\Delta_1^{(1)}\big) + (1-\omega^{(1)})\sigma_\beta\big(m\Delta_2^{(1)}\big)\big] + p_{\mathcal{B}}^{(2)}\big[\omega^{(2)}\sigma_\beta\big(m\Delta_1^{(2)}\big) + (1-\omega^{(2)})\sigma_\beta\big(m\Delta_2^{(2)}\big)\big]}{\big(1-p_{\mathcal{B}}^{(1)}\big)\big[\omega^{(1)}\sigma_\beta\big(m\Delta_1^{(1)}\big) + (1-\omega^{(1)})\sigma_\beta\big(m\Delta_2^{(1)}\big)\big] + \big(1-p_{\mathcal{B}}^{(2)}\big)\big[\omega^{(2)}\sigma_\beta\big(m\Delta_1^{(2)}\big) + (1-\omega^{(2)})\sigma_\beta\big(m\Delta_2^{(2)}\big)\big]}. \tag{B3}$$

If there are weights $\omega^{(g)} \in [0, 1]$ corresponding to the intersection point (D_1^*, D_2^*), then the strongly fragmented state exists. These states leave both markets equally active and as we showed in the discussion in Sec. V they entail benefits for the population as a whole, not favoring any of the symmetric groups.

Partially fragmented states. Finally, we outline a procedure to find a population steady state that is a combination of a bimodal (S) state in one group and a unimodal (U or W) state in the other, for $r \to 0$. A starting point for this search can be obtained by solving the homogeneous population equations (B2). When one of the groups is consistent with the homogeneous population assumption while the other is not, we can investigate whether the strongly fragmented solution for this other population exists. To find these states, we assume that the group that is inconsistent with a given homogeneous population solution is in the fragmented state. Thus possible order parameters for this state are on the locus defined by the Maxwell construction. For every pair (D_1, D_2) from the fragmented state locus we investigate the free energy of the second group (whether it is unfragmented or weakly fragmented). We find the peak position and represent the attraction distribution as a unimodal distribution centered at the (global) free energy minimum. We only need to examine whether by peak weight redistribution of the strongly fragmented group we can retrieve the initial order parameters (D_1, D_2). When this is possible, the partially fragmented state exists. In the example shown in Fig. 10, due to mild buy and sell preferences, when one of the groups is fragmented there are two unfragmented options for the second group, corresponding to specialization to either of the two markets.

[1] H. Mendelson, J. Financ. Quant. Anal. **22**, 189 (1987).
[2] B. Chowdhry and V. Nanda, Rev. Financ. Stud. **4**, 483 (1991).
[3] A. Madhavan, Rev. Financ. Stud. **8**, 579 (1995).
[4] B. Biais, L. Glosten, and C. Spatt, J. Financ. Mark. **8**, 217 (2005).
[5] P. Bennett and L. Wei, J. Financ. Mark. **9**, 49 (2006).

[6] H. Degryse, F. de Jong, and V. van Kervel, Rev. Financ. **19**, 1587 (2015).

[7] S. Buti, B. Rindi, and I. M. Werner, J. Financ. Econ. **124**, 244 (2017).

[8] C. Castellano, S. Fortunato, and V. Loreto, Rev. Mod. Phys. **81**, 591 (2009).

[9] J.-P. Bouchaud, Europhys. News **50**, 24 (2019).

[10] D. Helbing and P. Molnar, Phys. Rev. E **51**, 4282 (1995).

[11] M. Moussaïd, D. Helbing, and G. Theraulaz, Proc. Natl. Acad. Sci. USA **108**, 6884 (2011).

[12] G. Tedeschi, G. Iori, and M. Gallegati, J. Econ. Behav. Organ. **81**, 82 (2012).

[13] S. Galam, Y. Gefen, and Y. Shapir, J. Math. Sociol. **9**, 1 (1982).

[14] J. Fernández-Gracia, K. Suchecki, J. J. Ramasco, M. San Miguel, and V. M. Eguíluz, Phys. Rev. Lett. **112**, 158701 (2014).

[15] D. Challet, M. Marsili, and Y.-C. Zhang, Physica A **294**, 514 (2001).

[16] D. Challet and M. Marsili, Phys. Rev. E **68**, 036132 (2003).

[17] Z.-G. Huang, J.-Q. Zhang, J.-Q. Dong, L. Huang, and Y.-C. Lai, Sci. Rep. **2**, 703 (2012).

[18] M. Pagano, Q. J. Econ. **104**, 255 (1989).

[19] G. Ellison, D. Fudenberg, and M. Möbius, J. Eur. Econ. Assoc. **2**, 30 (2004).

[20] B. Shi, E. H. Gerding, P. Vytelingum, and N. R. Jennings, Autonomous Agents Multi-Agent Syst. **26**, 245 (2013).

[21] B. Caillaud and B. Jullien, RAND J. Econ. **34**, 309 (2003).

[22] G. Shorter and R. S. Miller, Dark pools in equity trading: Policy concerns and recent developments, Congressional Research Service Report No. R43739, Library of Congress, Washington, DC (2014).

[23] P. Gomber, S. Sagade, E. Theissen, M. C. Weber, and C. Westheide, J. Econ. Surv. **31**, 792 (2017).

[24] A. P. Kirman and N. J. Vriend, in *Interaction and Market Structure*, edited by D. Gatti, M. Gallegati, and A. Kirman, Lecture Notes in Economics and Mathematical Systems Vol. 484 (Springer, Berlin, 2000), pp. 33–56.

[25] K. Cai, E. Gerding, P. McBurney, J. Niu, S. Parsons, and S. Phelps, Overview of CAT: A market design competition, University of Liverpool, Department of Computer Science Report No. ULCS-09-005, 2009 (unpublished), Version 2.0.

[26] A. Alorić, P. Sollich, and P. McBurney, in *Advances in Artificial Economics*, edited by F. Amblard, F. J. Miguel, A. Blanchet, and B. Gaudou, Lecture Notes in Economics and Mathematical Systems Vol. 676 (Springer International, Cham, 2015), pp. 79–90.

[27] A. Alorić, P. Sollich, P. McBurney, and T. Galla, PLoS One **11**, e0154606 (2016).

[28] C. Camerer and T. H. Ho, Econometrica **67**, 827 (1999).

[29] T. H. Ho, C. Camerer, and J.-K. Chong, J. Econ. Theory **133**, 177 (2007).

[30] D. K. Gode and S. Sunder, J. Polit. Econ. **101**, 119 (1993).

[31] J. Duffy, Handb. Comput. Econ. **2**, 949 (2006).

[32] D. Ladley, Knowl. Eng. Rev. **27**, 273 (2012).

[33] M. Anufriev, J. Arifovic, J. Ledyard, and V. Panchenko, J. Evol. Econ. **23**, 539 (2013).

[34] N. Hanaki, A. Kirman, and M. Marsili, J. Econ. Behav. Organ. **77**, 382 (2011).

[35] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World* (Cambridge University Press, Cambridge, 2010).

[36] R. Nicole and P. Sollich, PLoS One **13**, e0196577 (2018).

[37] A. Alorić, Spontaneous segregation of adaptive agents in auctions, Ph.D. thesis, King's College London, 2017.

[38] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier Science, Amsterdam, 1992).

[39] K. Binder, Phys. Rev. Lett. **47**, 693 (1981).

[40] Y.-W. Cheung and D. Friedman, Games Econ. Behav. **19**, 46 (1997).

[41] J. Huber, J. Econ. Dyn. Control **31**, 2536 (2007).

[42] B. Toth, E. Scalas, J. Huber, and M. Kirchler, Eur. Phys. J. B **55**, 115 (2007).

RESEARCH ARTICLE

Forma mentis networks quantify crucial differences in STEM perception between students and experts

Massimo Stella[1,2]*, **Sarah de Nigris**[3], **Aleksandra Aloric**[4], **Cynthia S. Q. Siew**[5,6]

1 Institute for Complex Systems Simulation, University of Southampton, Southampton, United Kingdom, **2** Complex Science Consulting, Lecce, Italy, **3** Institute for Web Science and Technologies, University of Koblenz-Landau, Koblenz, Germany, **4** Scientific Computing Laboratory, Center for the Study of Complex Systems, Institute of Physics Belgrade, Belgrade, Serbia, **5** Department of Psychology, University of Warwick, Coventry, United Kingdom, **6** Department of Psychology, National University of Singapore, Singapore, Singapore

* massimo.stella@inbox.com

Citation: Stella M, de Nigris S, Aloric A, Siew CSQ (2019) Forma mentis networks quantify crucial differences in STEM perception between students and experts. PLoS ONE 14(10): e0222870. https://doi.org/10.1371/journal.pone.0222870

Editor: Diego Raphael Amancio, University of Sao Paulo, BRAZIL

Received: April 1, 2019

Accepted: September 9, 2019

Published: October 17, 2019

Data Availability Statement: The data is available on the Open Science Foundation public repository, under CC licence 4.0-By-Attribution, with the following DOI: 10.17605/OSF.IO/XYFWG.

Funding: M.S. was supported by an EPSRC Doctoral Training Centre grant (EP/G03690X/1). Official website: https://epsrc.ukri.org/. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Complex Science Consulting provided support in the form of salaries for author M.S., but did not have any additional role in the study design,

Abstract

In order to investigate how high school students and researchers perceive science-related (STEM) subjects, we introduce *forma mentis networks*. This framework models how people conceptually structure their stance, mindset or *forma mentis* toward a given topic. In this study, we build forma mentis networks revolving around STEM and based on psycholinguistic data, namely free associations of STEM concepts (i.e., which words are elicited first and associated by students/researchers reading “science”?) and their valence ratings concepts (i.e., is “science” perceived as positive, negative or neutral by students/researchers?). We construct separate networks for ($N_s = 159$) Italian high school students and ($N_r = 59$) interdisciplinary professionals and researchers in order to investigate how these groups differ in their conceptual knowledge and emotional perception of STEM. Our analysis of forma mentis networks at various scales indicate that, like researchers, students perceived “science” as a strongly positive entity. However, differently from researchers, students identified STEM subjects like “physics” and “mathematics” as negative and associated them with other negative STEM-related concepts. We call this surrounding of negative associations a negative *emotional aura*. Cross-validation with external datasets indicated that the negative emotional auras of physics, maths and statistics in the students’ forma mentis network related to science anxiety. Furthermore, considering the semantic associates of “mathematics” and “physics” revealed that negative auras may originate from a bleak, dry perception of the technical methodology and mnemonic tools taught in these subjects (e.g., calculus rules). Overall, our results underline the crucial importance of emphasizing nontechnical and applied aspects of STEM disciplines, beyond purely methodological teaching. The quantitative insights achieved through forma mentis networks highlight the necessity of establishing novel pedagogic and interdisciplinary links between science, its real-world complexity, and creativity in science learning in order to enhance the impact of STEM education, learning and outreach activities.

data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of this author are articulated in the 'author contributions' section.

Competing interests: Commercial affiliation with Complex Science Consulting does not alter the authors' adherence to PLOS ONE policies on sharing data and materials. All the authors declare no competing financial, non-financial, professional or personal competing interests.

Introduction

Increasing evidence indicates that many students develop a negative perception of STEM subjects before ending high school [1–3]. Mathematics is viewed as a difficult subject, physics is perceived as too abstract, and statistics is often considered an uninterpretable black box [3, 4]. A growing disinterest of students towards Science, Technology, Engineering and Mathematics (STEM) disciplines represents an unseen societal cost, as it translates into a lower interest in pursuing technological and scientific careers which are increasingly found to positively correlate with job growth, higher employment rates, societal innovation through functional literacy and economic development [5, 6]. Before addressing students' (mis)perception of STEM subjects, educators and policymakers first need to understand the detailed nature of the students' opinions and beliefs about science.

With this aim, this paper capitalizes on an innovative combination of methods from network science and cognitive science to examine the perception of STEM subjects among a population of students and another population of researchers. Specifically, we introduce the methodology of *forma mentis networks* (FMNs), which represent the associative structure of concepts as well as their valence, and show how FMNs can be harnessed to study a population's stance toward a given topic.

Forma mentis networks are constructed from language data. Linguistic information, such as text or speech, often conveys the opinion or attitude of an individual towards a given entity [7, 8], e.g., a human reading a blog can understand which posts are in favor or against a given political view. However, stance detection, i.e., detecting the stance of an individual or population from language [9, 10], is not an easy task.

In the past few decades, stance detection has spurred research at the interface of psycholinguistics and computer science, which has led to the development of a variety of methodologies through the human coding of grammatical features of text [9, 11], e.g. the use of specific adverbs or writing styles. Such approaches are centralised, in that they require a human coder, e.g. a linguist, to parse the input and detect the features that are important for the identification of the author's stance.

Centralised human coding cannot deal with the large volumes of linguistic data that are increasingly available, for instance from social media platforms [12]. This motivated the development of automatic techniques for detecting stance based on computer science approaches such as machine learning [10, 13, 14]. A notable example is a recent approach by Mohammad and colleagues [14], who deployed machine learning of sentiment features and word embeddings for successfully detecting the stance of individual messages from social media. The results of Mohammad and colleagues clearly show that stance detection is not the same as sentiment analysis. Sentiment analysis determines the specific affect valence of a given piece of linguistic data, i.e. how universally positive/negative/neutral are the concepts elicited by a given portion of text. Instead, stance emerges at a higher level as a non-trivial combination of different patterns of affect and sentiment. For example, the sentence "The dictator who killed my relatives has been finally executed" includes concepts of negative sentiment (e.g. dictator, executions, etc.), but nonetheless elicits a positive stance towards the execution itself. It is important to underline the additional complexity of stance in comparison with sentiment, as affective patterns in the language need to be integrated with additional contextual information before achieving an accurate classification of stance itself [14].

Although machine learning approaches are powerful in underlining the different psychological dimensions of stance in terms of context and sentiment [14], these automatic techniques have at least two limitations: (i) performance depends on the availability and quality of large-scale annotated training data, and (ii) machine learning builds "black-box" representations of data that cannot be directly accessed or interpreted. Due to these two elements, supervised learning approaches to stance detection are not yet widespread in the cognitive sciences, although they represent an interesting and powerful perspective for future work.

Beyond supervised learning, network models stand as a promising avenue to the investigation of cognitive and linguistic data, leading to the emergence of the field of cognitive network science [15]. Network models of language are often interpreted as descriptive representations of the *mental lexicon*, a repository of linguistic and semantic knowledge in human memory [7]. Decades of research in psycholinguistics has shown that the mental lexicon is not a static list of words, e.g. a dictionary, but it rather is a dynamical system optimized for cognitive computing which stores and processes individual concepts together with their associated linguistic data, e.g. semantic overlap in meaning [16], phonological similarities [17, 18], syntactic relationships between word categories [19]. Psycholinguistic evidence has shown that the associative structure of the mental lexicon influences language processes such as word learning [20–22] and processing [16, 23–25]. This strong link between mental lexicon structure and language usage promoted the use of network models for a variety of processes such as the discovery of writing styles and text authorship from word co-occurrences in texts [26, 27], improving the accuracy of clinical diagnosis of Alzheimer's Disease risk [28], modelling and understanding the success rates of picture naming in people with aphasia [29], predicting the creativity of individuals [30–32], their curiosity [33, 34], their openness to new experience [35], their expertise in a given domain [36, 37] and their perceived anxiety toward a topic [38]. Forma mentis networks rely on the framework of cognitive network science to represent the associative and emotional structure of concepts in the mental lexicon.

One of the main ingredients of FMNs is free association data to specify the connections between concepts. Indeed, free associations represent a powerful and meaningful way of building network models of the mental lexicon [23–25, 31]. Free associations are obtained empirically from experiments where participants have to produce associates when primed with a cue word. Hence, free associations are largely free from any specific semantic definition (e.g., synonyms). Previous work [19, 20] has shown that free associations partially overlap with other semantic word-word similarities such as synonyms (i.e., two words sharing the same meaning in a given context) or generalisations (i.e., a concept being a special type of another word) but also display a small overlap with phonological similarities among words (e.g., when pronunciations differ in one phoneme).

Forma mentis networks combine free associations with affective patterns of concepts. In a FMN, nodes represent concepts or words, links indicate free associations provided by a given population and every node has a valence attribute [39–41] that represent how the population perceives a given concept or word (positive, negative, or neutral). Recent psycholinguistic evidence has shown that the emotional valence of words influences language processing and memory [41, 42], highlighting an important link between affect and the cognitive mechanisms of language processing in the mental lexicon. Therefore, representing knowledge and sentiment combined in a forma mentis network gives access to the structure of the aggregated mental lexicon and affect of a given population.

We emphasize that the addition of valence attributes and the adoption of free associations makes forma mentis networks different from conceptual maps [43–46], which represent important network models of knowledge acquisition and structuring during learning but do not incorporate information about how learners perceive individual conceptual units. Another difference is that conceptual maps are often based on concept co-occurrence in a syllabus and therefore capture temporal information [46], which is not present in a forma mentis network.

Notice that forma mentis networks rely on free associations, which capture the associative structure of semantic memory [23–25] through an empirical assessment of which concepts quickly remind of each other. Hence, these associations mirror memory patterns and are "free" from basic communication demands in sentences, such as the need to link words according to specific syntactic rules [25]. This aspect makes free associations qualitatively different from other types of word-word relationships like word co-occurrences or syntactic dependencies [27] which rather capture syntactic relationships (e.g., a verb being related to a noun). In quantitative terms, co-occurrences and syntactic dependencies can be derived automatically from written corpora, whereas free associations usually require a behavioral experiment. Furthermore, it is important to underline that free associations might overlap with syntactic dependencies but also comprise a wider variety of conceptual associations [7, 23, 25], ranging from sound similarities to meaning overlap, from visual similarity to semantic feature sharing. Forma mentis networks build on this richness of associative knowledge for representing the mindset or *forma mentis* of groups of individuals.

In this paper, we investigated the attitude of students and researchers towards science and STEM subjects through this innovative combination of tools from psycholinguistics and network science to quantify their stances toward STEM subjects. Through the comparison of the FMN of students and research professionals, we provide quantitative evidence for sharp differences in the perception of STEM among the two different groups. Specifically, the combination of conceptual associations and affect patterns allowed us to identify and paint a richer picture of the disaffection towards mathematics and physics exhibited by students and absent in STEM professionals.

Methods

Participants

We collected data from 159 students and 59 researchers. Students were selected from three different Italian high schools, without consideration of their grades in STEM subjects. All students were in their final year of high school, with ages ranging between 18 and 19 years (mode: 18 years). In order to build a sample representative of the national Italian student population in high schools, entire classes were selected for testing, to ensure a mixture of socio-economic backgrounds and STEM proficiency levels. Participants were roughly evenly distributed between female (53%) and male (47%) students.

Researchers were selected from large-scale international workshops because of the necessity of physically interviewing large numbers of experts at once. Our selection focused on early career scientists, which included doctoral students and post-doctoral researchers, with the aim of including as many diverse backgrounds as possible. We focused our selection towards researchers applying quantitative tools originating in the fields of mathematics, physics and computer science to study emerging phenomena in complex systems ranging from biological to socioeconomic systems. Hence, all the interviewed researchers possessed advanced training and expertise in STEM and were actively pursuing a professional career in science. The age of the interviewed researchers ranged between 24 and 39 years, with a mode of 29 years. Participants were roughly evenly distributed between male (56%) and female (44%) researchers.

Cognitive tasks

Each participant took part in a survey composed of two tasks: (i) a free association task and (ii) a valence evaluation task. Participants were given precise instructions about the study before proceeding. Participants were then asked to provide informed consent if they agreed to take part in the study by signing a consent form that described essential points about privacy and ethics. All consent forms were gathered at the end of the study and are available upon inquiry to the first author.

In the free association task, each participant was presented with a list of 50 cue words. In order to investigate attitudes toward key STEM subjects, 10 out of the 50 cue words were present in all participants' lists. These words were: *mathematics*, *complex*, *physics*, *chemistry*, *system*, *biology*, *life*, *art*, *school* and *university*. In Italian, these words were translated as: *matematica*, *complesso*, *fisica*, *chimica*, *sistema*, *biologia*, *vita*, *arte*, *scuola* and *università*. Although "art" is not a STEM subject, its inclusion in the list of essential words was meant to provide some comparison between the humanities and technical subjects. Additionally, adding art to the list of essential words provides a way to probe students' perception of connections between STEM subjects and creativity, a link that has been investigated in previous studies about attitudes towards STEM [3]. The other 40 cue words were drawn at random from a subsample of STEM-focused 390 words. The pool of 390 potential cue words was obtained by considering the highest frequency non-stop words from the Wikipedia webpages about "Complex System", "Physics", "Mathematics", "Biology", "Chemistry" and "Psychology" (as accessed on: 15 January 2017).

Participants were randomly assigned to one of the 50 files containing a different random realisation with 50 of the previously described words. The order of words in each list was scrambled with the aim of reducing recency effects or other associative biases due to the order of cues.

In order to obtain denser networks of free associations, we used the continuous free association task, which has been shown to provide higher quality data that could account for more variance in lexical retrieval tasks [23]. In the continuous free association task, each participant generated three associative responses to each item in the list of 50 cue words. The association task took place in a lab setting, with each participant filling in electronic forms on a computer terminal while under supervision. Forms that contained more than 25 percent blank responses were discarded. This occurred in roughly 2% of the cases. The first three associates in each form were discarded in order to minimize potential priming effects that might follow the given instructions. The association task lasted for 10 minutes, followed by a short break.

In the second task, we collected the valence of each cue and for all associations from the free association task. Participants were asked to rate the valence of each cue and their associated responses using a Likert scale ranging from 1 (very negative) to 5 (very positive), with neutrality being represented either by a blank space or by a score of 3. Participants completed this task in about 10 minutes. Those participants who did not finish the task in 10 minutes left the remaining spaces blank. Forms that contained more than 25 percent blank responses were discarded. This occurred in roughly 3% of the cases. Data collection was conducted anonymously, such that no demographic or educational data was obtained from the participants and directly linked to the filled forms.

Data cleaning and network building

Associative responses were converted to lowercase letters and checked automatically and manually for common spelling mistakes. The automatic spell checkers used were based on Google Translate and Wolfram's Mathematica 11.3 (manufactured by Wolfram Research, Champaign, US). Different word forms were manually converted to match their singular forms (e.g. in English "muscles" was changed to "muscle") and composite responses were changed to single-word forms (e.g. in Italian "da dove" was changed to "dove").

A forma mentis network was constructed such that nodes represented lexical items and edges indicated free associations between words. Two networks of free associations were constructed from the students' data, one network where only associations provided by at least two different participants were considered $\mathcal{N}_S^C$ (filtered) and a second network where no filtering was performed $\mathcal{N}_S$. Given the considerably smaller sample size of researchers, only a single unfiltered network of free associations was constructed $\mathcal{N}_R$. We note that considering idiosyncratic associations, i.e., associations provided by a single participant, like in $\mathcal{N}_S$, is common practice when working with free association data obtained from small samples [43, 44, 47] and they are still considered to be insightful of cognitive patterns [48]. Each node in the network was also assigned a valence score and an attribute ("positive", "neutral", or "negative").

In the remainder of the paper, *forma mentis network* refers to the network representation that simultaneously represents conceptual knowledge derived from free associations and the emotional perceptions of those concepts among a given population (i.e., students or researchers).

Statistical analysis of word valence

In order to categorize positive, neutral and negative concepts we used a non-parametric statistical test (Kruskall-Wallis test). The statistical test was used to assess whether the scores attributed to word *i*, namely w_i, had a lower, compatible, or higher median valence as compared to the remaining distribution of valence scores, in formulas $\bigcup_{j \neq i} w_j$. Non-parametric testing was used because the original distribution of valence scores $\bigcup_j w_j$ were skewed with a heavy left tail (Pearson's skewness coefficient $s_s = 3(mean_s - median_s)/\sigma = 1.39$ for students' data and $s_r = 1.45$ for researchers). Concepts which had a median valence score lower than the rest, according to a Kruskall-Wallis test with significance level $\alpha = 0.1$, were labelled as *negative*. Concepts which had a median valence score higher than the rest, according to a Kruskall-Wallis test with significance level $\alpha = 0.1$, were labelled as *positive*. Remaining concepts were labelled as neutral.

Defining valence beyond lexical items: Valence "auras"

Valence is a commonly used feature in psycholinguistic models that assesses the sentiment of a given text [8, 39]. At the word level, the valence of a word represents the positive, neutral, or negative connotation is elicited by the word in a given population. Hence, valence is a feature of individual words, and does not further consider the way in which words are associated with each other.

Combining free associations with their valence in forma mentis networks naturally provides a way of extending the concept of word valence to a given cluster of associated words. We introduce the concept of *valence aura*, which identifies the valence of the immediate neighbors of a word on a network of free associations. A concept has a negative valence aura if the given concept is associated with more negative concepts than it is associated with positive concepts. On the other hand, a concept has a positive valence aura if the given concept is associated with more positive concepts than it is associated with negative concepts. The polarity of an aura is determined by the most frequent valence of words in the neighborhood (following a majority rule). It is important to note that positively valenced words could have either a negative or positive valence aura, and negatively valenced words could have either a negative or positive valence aura. Hence, valence auras provide us with a way of juxtaposing or contrasting sentiment polarities with consideration of the structural organisation of knowledge beyond how individual concepts are valenced in isolation.

We use the methodology of valence auras to investigate potential differences in the way that students and researchers structure their conceptual knowledge and the role of the perceived valence of those concepts. Firstly, within a given population, it would be interesting to assess the tendency of positive concepts to be associated with other positive concepts (i.e., to have a positive valence aura), and equivalently for negative concepts (i.e. negative concepts with other negative ones). This would bolster the idea that sentiment polarities of individual words have the potential of influencing the structural organisation of semantic memory. Secondly, and more importantly, there might be different tendencies to surround positive concepts with positive or negative auras between students and researchers. Differences in the mixing of (individual) word valence and word auras between the forma mentis networks of different populations could potentially highlight important differences in the organisation and perception of knowledge between such populations.

Validation of the definition of auras through additional psycholinguistic data

In order to validate our operationalization of valence auras and the valence ratings collected in the present study, we used external datasets of word valence for comparison. For English words in the researchers' forma mentis network, we used the affective ratings by Warriner and colleagues [40], whereas for Italian words in the students' forma mentis network we used the valence norms recently gathered by Fairfield and colleagues [49]. Kendall tau correlation tests indicate that in both cases there is a statistically significant positive correlation (at $\alpha = 0.1$) between the mean valence scores collected in the current study and the ones obtained from previous investigations (for students: $\tau = 0.51$, $p < 10^{-5}$; for researchers: $\tau = 0.38$, $p < 10^{-5}$). For students, the dataset by Fairfield et al. contained valence scores for only 491 of the 4483 words in the Italian unfiltered forma mentis network. For researchers, the overlap between our dataset and Warriner and colleagues' was higher, covering 1173 of the 1616 words in the English unfiltered forma mentis network.

In the following, we will not use directly the mean valence scores gathered from students or researchers but rather valence attributes (i.e., positive, neutral, or negative) to define the valence auras of words. In the results section, we will show that our retrieved valence attributes are compatible with the valence scores obtained by other studies.

Results

Network structure and valence identify auras, which in turn identify words of extreme valence and arousal

The operationalization of valence auras as reported in the Methods section combines network structure with the valence of individual words. Does this combination of topological and valence information provide further insights into students and researchers' perception of STEM subjects? In order to answer this question, we compared the mean valence and arousal scores from external psycholinguistic datasets (cf. Methods) of negative words with either positive or negative auras, and positive words with either positive or negative auras. Recall that auras were defined by using the valence ratings and network structures obtained from the cognitive tasks described above, whereas mean valence scores come from external sources (i.e., the Fairfield [49] and Warriner [40] databases).

Fig 1 reports the mean valence and arousal of words from the students' forma mentis network. In the students' FMN, negative words surrounded by a negative aura have a lower mean valence (based on an external dataset, cf. Methods) as compared to negative words surrounded by any aura. This difference was statistically significant at the 0.1 significance level α (Kruskal-Wallis, $N = 116$, $s = 2.8994$, $p = 0.088$). The difference in mean valence between positive words surrounded by any aura and positive words surrounded by a positive aura was not statistically significant at the $\alpha = 0.1$ level (Kruskal-Wallis, $N = 238$, $s = 2.2891$, $p = 0.1314$). The forma mentis network of students also highlighted an interesting difference in the mean arousal of negative

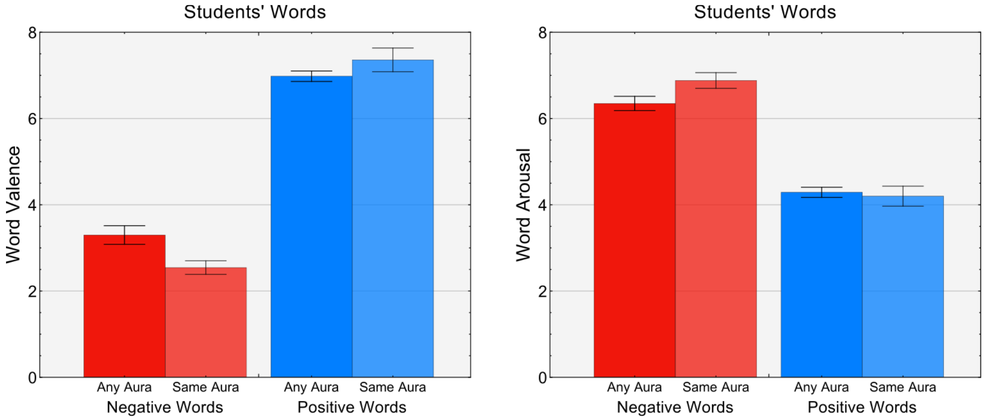

Fig 1. Valence auras identify more extreme negative words in the students' population. Mean valence (left plot) from external psycholinguistic datasets of negative/positive words either surrounded by any valence auras (full colours) or by the same aura (lighter colours). In the students' forma mentis network, negative words surrounded by negative auras are perceived as more strongly negative on average as compared to negative words surrounded by any aura. Negative words surrounded by a negative aura had higher arousal than negative words with any aura (right plot). No difference was found for positive words.

https://doi.org/10.1371/journal.pone.0222870.g001

words surrounded by a negative aura as compared to negative words surrounded by any aura (cf. Fig 1, left plot). Negative words surrounded by a negative aura elicited a stronger arousal as compared to negative words surrounded by any aura, and the difference was statistically significant at the $\alpha = 0.1$ level (Kruskal-Wallis, $N = 116$, $s = 2.7338$, $p = 0.0984$). A threshold of 0.1 was chosen in order to cope with the limited overlap of our data with the external databases.

The above differences suggest that negative auras correspond to a boosted arousal when surrounding words of negative valence. Positive words did not elicit any analogous difference. No statistically significant difference was found among words in the researchers' forma mentis network.

These results indicate that in the organisation of STEM-related concepts as represented by the forma mentis network, negative concepts surrounded by a negative aura are in general perceived as more negative and elicit higher arousal than concepts surrounded by any aura. Since the arousal scores might be a by-product of valence in rating experiments [41], negative words surrounded by a negative aura represent negative concepts that can activate or be activated by other negative concepts and thus lead to an increase in arousal and emotional intensity. The above analysis presents quantitative evidence showing that our operationalization of valence auras of an individual word's direct neighbors in an associative network can highlight additional affective patterns that valence scores of individual words cannot.

The unfiltered forma mentis network of researchers is smaller, less connected and has fewer negative words than the students' forma mentis network. Hence, no significant differences were found at the $\alpha = 0.1$ level of significance (Kruskal Wallis, for positive words: $s = 1.0674$, $p = 0.3033$; for negative words: $s = 0.0574$, $p = 0.8119$). We hypothesized that this is due to the forma mentis network of researchers lacking the resolution or power to detect distinct affective patterns.

Forma mentis networks show assortative mixing of word valence

As described above, in a forma mentis network, each word has a valence attribute (e.g. "positive" or "negative"). Links represent associations between two concepts but also between their respective valence attributes (e.g., there can be a link connecting a "positive" concept to a "negative" concept). Investigating the assortative or disassortative mixing of valence attributes across network links can shed light on potential trends in students' and researchers' structural and emotional organisation of knowledge.

In the unfiltered network of student associations $\mathcal{N}_S$, a Kendall Tau test between the valence attributes of links' endpoints reveals a statistically significant positive correlation $\tau = 0.163$, $p < 10^{-5}$. This value might indicate that students tend to associate positive (negative) concepts to other positive (negative) concepts. However, comparison with a reference null model is necessary in order to assess the relative strength of the above correlation and test whether it might be a direct consequence of either the distribution of node degree or the counts of positive, neutral, or negative attributes. We used as null reference a configuration model fixing both the empirical degree distribution and the valence attributes of words in the original network but randomising links. An average Kendall Tau of $\tau_r = -0.0001$ ($p > 0.310$) was obtained over 50 independent realisations of the null model. As the empirical correlation $\tau = 0.163$ was several orders of magnitude larger than random expectation, this indicated that there was a strong tendency for students to associate positive (negative) concepts with other positive (negative) concepts independently of the distributions of either degree or valence attributes. We found a similar pattern in the way that researchers organised and perceived their STEM knowledge (empirical Kendal Tau $\tau = 0.116$, $p < 10^{-5}$, reference null model $\tau_r = 0.027$, $p > 0.112$).

Forma mentis networks indicate clustering of word valence and valence auras

In this section we investigated whether there was a tendency for words to be surrounded by other words with the same valence. To do this, we attributed arbitrary scores to valence attributes, i.e. $-m$ to negative words, 0 to neutral words and $+m$ to positive words. We used $m = 1$ for convenience, although our correlation analysis does not depend on the specific value of m.

In the students' forma mentis network $\mathcal{N}_S$, a Kendall Tau test between the valence attributes of a word and the average valence attributes of its neighbors revealed a statistically significant positive correlation $\tau = 0.385$, $p < 10^{-5}$. With 50 independent realisations of a configuration null model with fixed word attributes and degrees but random associations, we found an average correlation of $\tau_r = 0.053$ ($p = 0.060$). A similar result was found for the researchers' forma mentis network $\mathcal{N}_R$, where the empirical correlation value ($\tau = 0.323$, $p < 10^{-5}$) was considerably higher than random expectation ($\tau_r = 0.060$,$p = 0.056$).

Given that the empirical correlations were several orders of magnitude larger than the reference values, our results showed a tendency for both students and researchers to associate words of a given valence with auras of the same valence, i.e., negative concepts tend to have a negative aura whereas positive concepts tend to have a positive aura. This indicates that the forma mentis networks of researchers and students are on average highly clustered in neighborhoods of words with similar valence attributes. Deviations from this general trend, such as negative words in the positive aura of a positive word, can be informative of the way in which a given population perceives STEM subjects.

Forma Mentis networks highlight differing stances towards STEM subjects

In this section we focus our attention on the semantic content of associations. Fig 2 reports the attribute and aura of the 10 words that were always provided to participants as cues (see Methods). Positive (negative) words are highlighted in cerulean (red) for both students (top panel) and researchers (lower panel). Researchers associated STEM-related words with mainly positive concepts, whereas students associated STEM-related words with both positive and negative concepts.

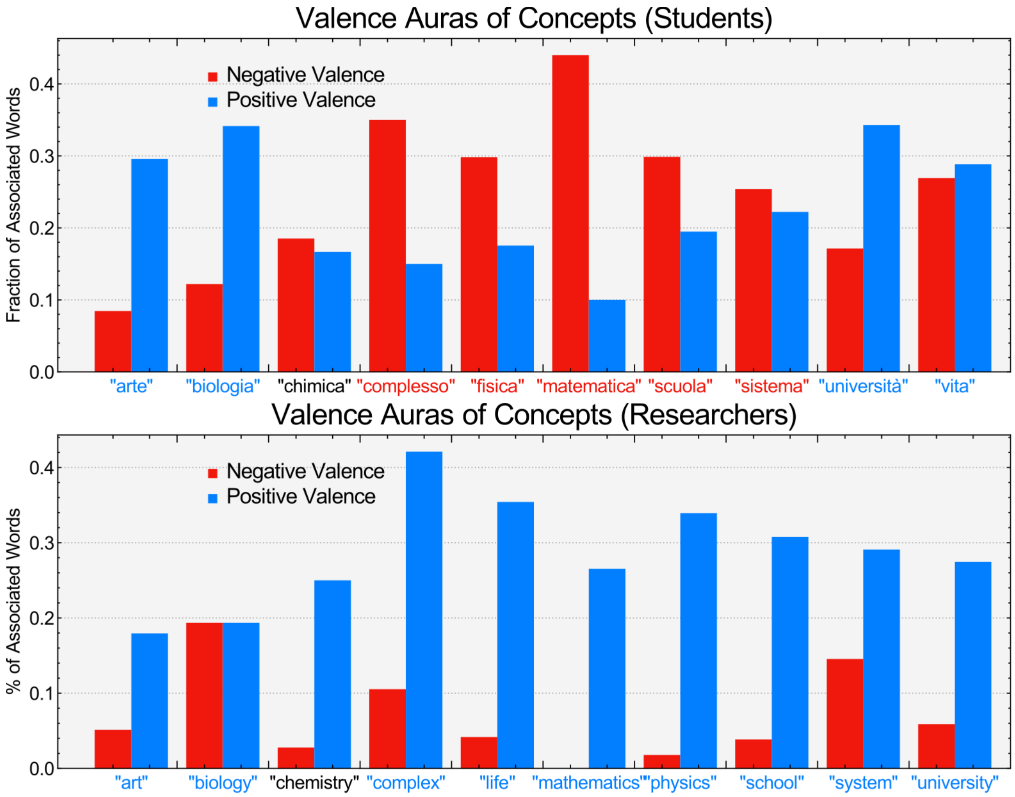

Fig 2. Valence auras identify a negative stance of students towards specific STEM subjects. Top panel: Fraction of neighbors of concepts having a positive or a negative average valence. Concepts with positive valence are reported in blue. Concepts with negative valence are reported in red. Students perceive quantitative scientific subjects such as Mathematics and Physics negatively. Students also attributed a negative aura to these disciplines, i.e., they associated Physics mainly with other negative, rather than positive, concepts. The auras of negativity were not aimed towards all STEM subjects, since biology was perceived as positive and surrounded by an aura of positive valence. **Bottom panel**: Valence auras for researchers. Notice that all essential concepts were perceived as positive and surrounded by auras of positive valence.

https://doi.org/10.1371/journal.pone.0222870.g002

The analysis of individual words reveals that researchers perceived almost all the 10 STEM-related words as positive concepts. On the other hand, students perceived words such as “complex”, “physics”, “mathematics”, “school” and “system” as negative concepts. Furthermore, the network structure of forma mentis networks highlighted additional critical differences in the way that students and researchers attributed positive or negative auras of valence to such negatively perceived STEM words. Specifically, students associated concepts such as “physics” and “mathematics” to other negative concepts, surrounding STEM words of quantitative disciplines with a negative valence aura. The ratio of negative to positive concepts is particularly high in the case of “mathematics”, where almost 43% of associations were to other negative concepts. These patterns were absent in the forma mentis network of researchers. This comparison suggests that the presence of negative auras attributed to some STEM-related concepts is not merely a consequence of the network construction but reflects the negative stance that students have towards quantitative disciplines such as physics and mathematics.

However, it is worth noticing how the data also indicates that students did not perceive all STEM subjects as negative. In fact, concepts such as “biology” and “university” were perceived as positive and were surrounded by a positive valence aura in the students’ forma mentis network. This contrast suggests that the aversion of STEM-related concepts might be related to

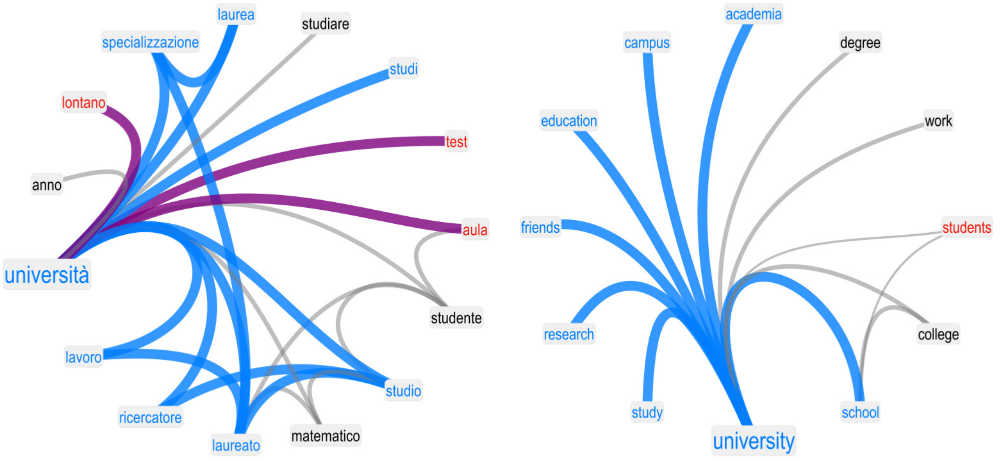

Fig 3. Neighborhoods in forma mentis networks determine the valence aura of concepts. Examples of the forma mentis networks in the neighborhood of "university" for students (left) and researchers (right). In a forma mentis network, nodes have valence attributes, i.e. "positive" (cerulean), "neutral" (grey) and "negative" (red). Links are weighted based on the number of participants providing a given association between concepts. Links between positive (negative, neutral, opposing) concepts are cerulean (red, grey, purple). The above examples include only associations provided by at least two participants. The neighbors surrounding a given word, together with their valence attributes, constitute the valence aura of that word. Both students and researchers perceive "university" as a positive concept and surround it with a positive aura.

https://doi.org/10.1371/journal.pone.0222870.g003

the quantitative disciplines that underlie the scientific method used in these fields. Fig 3 shows the neighborhoods of "university" in the filtered forma mentis network of students (left) and researchers (right). Notice that "university" is perceived as positive and surrounded by other positive concepts such as "degree", "study", "work", and "specialisation". Even the word "researcher" is positively perceived by students, indicating a positive stance towards the general concepts of research and education. Importantly, students strongly associated the concepts of "studying" and "work", as indicated by the presence of this connection in the statistically filtered FMN. This result suggests that students might be aware of the positive impact of education has with respect to future success in the job market [5, 6].

Fig 4 shows the neighborhoods of "physics" (top) and "mathematics" (bottom) in the filtered forma mentis network of students (left) and the FMN of researchers (right). Notice how concepts such as "physics" or "mathematics" gave rise to mostly negative associations in the students' population. The hierarchical edge bundling visualisation implemented in Mathematica 11.3 highlights that negative associations tend to cluster together, in agreement with the above clustering analysis. An inspection of the semantic content of the neighborhood for "mathematics" reveals the presence of clusters of negative concepts associated with the topic of calculus and geometry. Interestingly, most of these negative concepts were concrete tools and methodologies used in mathematics (e.g. "algorithm", "derivative", "graph", "theorem"), rather than abstract, more general terms such as "complexity". A similar result holds for "physics" (e.g., "function", "test", "integral"). A closer look at the semantic information embedded in the negative aura surrounding "physics" and "mathematics" provides preliminary evidence that the negative perception students have of these subjects may come predominantly from a negative perception of the quantitative tools usually taught in schools. In other words, the negative aura surrounding "mathematics" or "physics" does not come from a general negativity towards the whole educational system but rather from specific, concrete elements of teaching curricula. Improving the appreciation of students towards these concrete tools (e.g. "algorithm", "graph", "function") might have a beneficial effect on

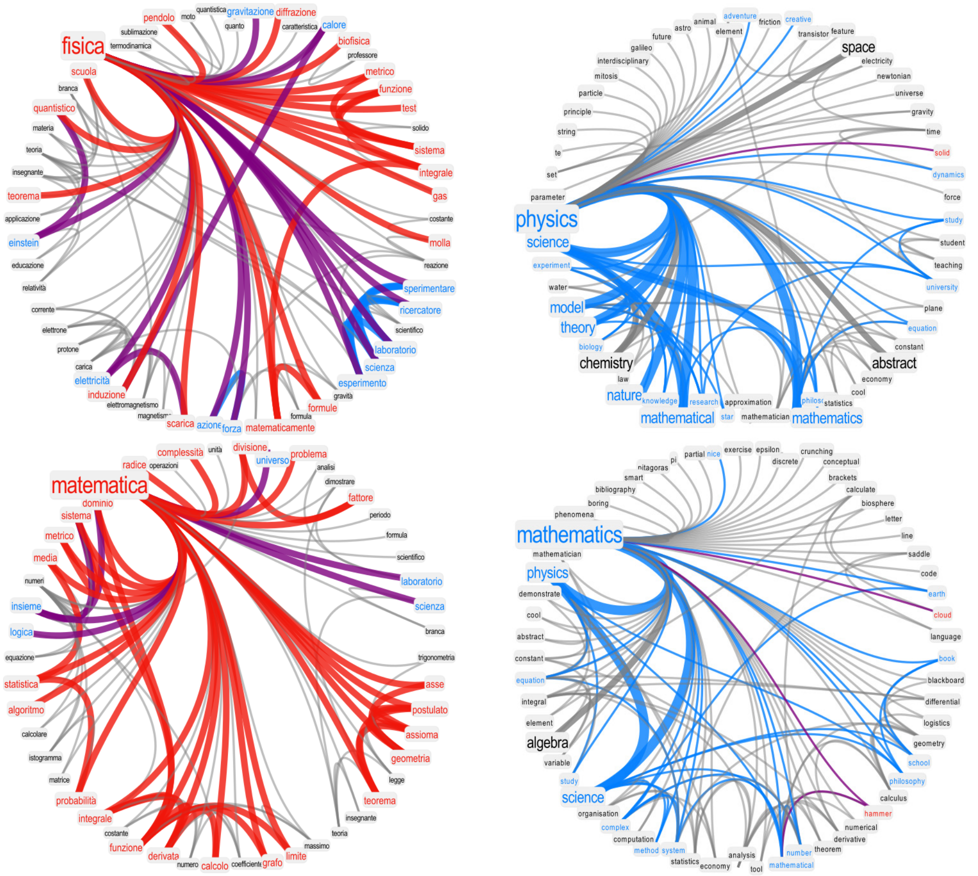

Fig 4. Mathematics and physics are perceived differently by students and researchers. The neighborhoods of "physics" (top) and "mathematics" (bottom) for students (left) and researchers (right). Red links indicate associations between concepts of negative valence. Stronger, more frequent associations are thicker. Students not only perceive "mathematics" and "physics" as negative concepts but also surround them with strongly negative auras of valence. This phenomenon is absent in the forma mentis network of researchers, indicating a critical negative attitude of students towards STEM quantitative subjects. Notice that for both physics and maths in students the negative aura comes mainly from clusters of specific concepts relating to specific tools (e.g. "derivative", "test", "integral").

https://doi.org/10.1371/journal.pone.0222870.g004

the perception that students have of "mathematics" or "physics", given the average trend reported above indicating that positive concepts tend to be surrounded by positive auras. However, it is important to note that our study by itself cannot either prove or disprove a causal link about concepts being perceived as positive because of their positive auras and additional research is required.

However, it is important to underline that although "mathematics" and "physics" displayed negative auras, in both the unfiltered and filtered forma mentis networks students were able to associate these subjects to "science", which is perceived as a positive concept (see also Fig 5). This link may indicate that students are aware of the importance of quantitative disciplines for the advancement of science.

Forma mentis networks further highlight the critical negative perception of students towards "mathematics" and "physics". As reported in Fig 5, those are the only negative concepts in the otherwise positive valence aura surrounding "science". This contrast indicates that, although on a technical level students were aware about the links between quantitative disciplines and science, they were unable to transfer their positive perception of science to the building blocks of the scientific method.

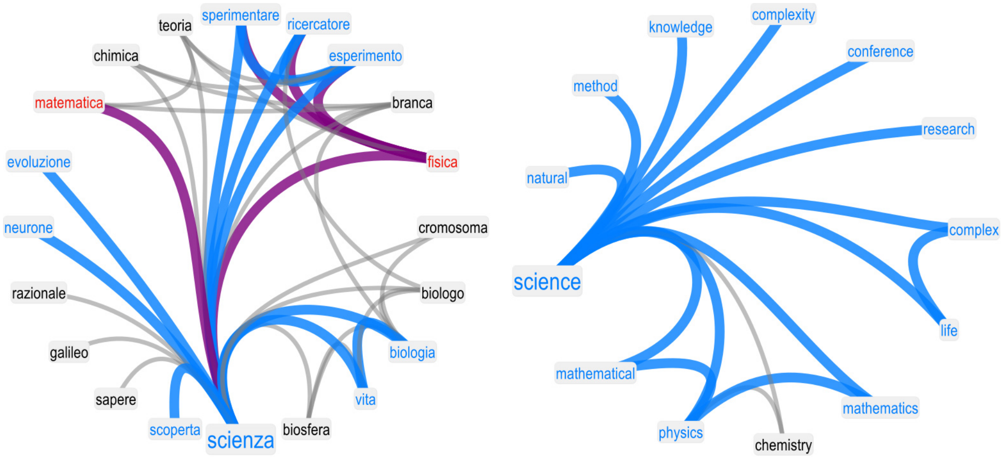

Fig 5. Maths and physics are the main negative outliers in the otherwise positive aura of "science" as perceived by students. "Science" was never provided as a cue either to students or to researchers. It was one of the associations provided by participants. Neighborhoods of "science" in the forma mentis networks of students (left) and researchers (right): although students perceived science and other STEM subjects as positive concepts, surrounding science itself with an aura of positive valence, students also perceived mathematics and physics as negative concepts. This plot included only statistically significant free associations.

https://doi.org/10.1371/journal.pone.0222870.g005

Discussion

Forma mentis networks represent an innovative combination of free associations, i.e. words that are elicited in response to cue words [48], with additional affective information of each word's valence [39, 40], i.e. how positively or negatively a given concept is perceived.

From a methodological perspective, this combination of two sources of linguistic information fills a gap in the literature of language networks modelling the mental lexicon [15], where conceptual units are considered only in terms of their semantic features [16, 23, 25, 30, 31], and not their valence. Given that recent evidence indicates that emotions deeply influence language processing and memory even at nonconscious levels [41, 42], forma mentis networks represent a natural extension of semantic representations of the mental lexicon that includes affective attributes of individual concepts.

The combination of sentiment and semantic structure leads to the definition of valence auras, in which concepts are not isolated affective entities but rather interacting emotional elements of an associative network. The empirical evidence reported in this work shows a tendency for concepts of a given valence to cluster with words of the same valence. This assortative mixing of links is known as homophily in social network analysis [50] and represents a tendency for units to link mainly to other units sharing similar features. To the best of our knowledge, our work represents the first evidence of *emotional homophily* in the human mental lexicon. This emotional homophily leads to concepts of a given valence being surrounded by other concepts with the same valence. By cross-validating our data with independent datasets of affective norms [40, 49], we showed that negative words surrounded by a negative aura elicited a higher arousal compared to negative words surrounded by any (neutral, negative, or positive) aura. This difference could be interpreted in terms of individual concepts exerting an influence over their associated neighbors, with negative words increasing levels of emotional intensity elicited during cognitive processing such that negative words surrounded by other negative words have higher arousal ratings than expected. This interaction between emotional processing and the structure of semantic memory itself should be further explored in future psycholinguistic research.

In the present paper, emotional homophily in forma mentis networks provides us with new ways of detecting the stance of a given population. Our application of the forma mentis networks has led to three main insights into the way that students and researchers perceive STEM topics.

First, we used the network structure of FMNs to identify and define the "aura" of a concept, i.e., its first neighbors in the association network created by the participants. An analysis of valence auras, in addition to the words' individual valence attributes, uncovers a clear pattern in the students' FMN in which negative words surrounded by a negative aura were correlated with higher arousal ratings. Moreover, words of the same valence tended to cluster together, indicating a conceptual organization that may have been shaped by the valences of words. Overall, it appeared that the students' stance towards STEM subjects is mixed, combining both positive and negative stances, whereas the researchers' stance toward STEM was predominantly positive.

Second, at the semantic level, the comparison between students' and researchers' networks led both to unexpected similarities as well as interesting contrasts. On the one hand, Italian high school students are almost as skilled as experts in relating key concepts of STEM subjects to "science", such that "science" itself was a key concept in their forma mentis network. This finding provides evidence that at a global level Italian students possess a good technical awareness of STEM subjects in comparison with STEM professionals. Analogously, a comparable level of student competence has also been reported in other educational systems, such as the Finnish one, by independent studies from Koponen and Nousiainen using concept maps [46].

Furthermore, students perceived concepts such as "mathematics" and "physics" not only as negative, but surrounded by negative auras as well, whereas words such as "science" were positively perceived by both students and researchers. This dichotomy between the positive aura of science and the negative auras of mathematics and physics is absent in the group of STEM professionals, and it suggests that students might not be sufficiently aware of the connections between science, its methods and its applications. Another interpretation of the negative auras surrounding mathematics, physics and other concepts like statistics (cf. S1 Appendix) might relate to emotional homophily and anxiety. As discussed above, negative concepts surrounded by a negative aura tended to also have higher arousal and lower valence ratings (based on external datasets). In the circumplex model of emotions [39], higher arousal and negative valence correspond to emotions of stress and anxiety. Hence, the negative emotional auras of physics, maths and statistics in the forma mentis of students suggest that high school students may experience stress and anxiety toward such disciplines. This finding is supported by an increasingly developing literature about mathematics anxiety [4], physics anxiety [51] and even statistics anxiety [38] affecting students' learning at the high school level and continuing even through university. Notice that emotional auras and anxiety represent a crucial problem in Education, since recent results show that stress and anxiety inhibit the acquisition and retention of STEM-related concepts [52]. When interpreted against the relevant literature, our results concretely point out the urgency for identifying and acting upon negative auras/perceptions in student populations in order to enhance STEM learning. Forma mentis networks represent an innovative way of quantifying science anxiety in student populations, potentially working in synergy with other network studies quantifying anxiety levels with complex networks [38] and psychological methodologies [4].

In spite of the bleak perception of mathematics and physics, the students showed a positive perception of both "university" and "researcher", which indicates an awareness of the importance of education for their professional future.

Third, valence auras allowed us to hypothesize viable reasons for the disaffection towards mathematics and physics beyond anxiety: Words with a negative valence in those subjects' aura are mostly mathematical tools and techniques, such as "integral" or "function". This result highlights how the negative stance on physics and mathematics might not originate from a wider distrust of the subjects themselves but perhaps from the difficulty in seeing the value of these techniques, particularly when devoid of interdisciplinary connections. This result opens possible avenues for intervention in education by, for instance, helping students to embed mathematics and physics within a richer network of conceptual associations.

What might be missing from the educational curriculum of the students participating in this study is an emphasis on the connections between quantitative disciplines and real-world settings. Beyond sterile arguments that a discipline should be appreciated because of its inner beauty, our results suggest that even at the high school level, educators should provide as many opportunities as possible for students to discover the beauty of STEM subjects and learn about the implications of mathematical modelling of real-world systems, as previously highlighted in the relevant educational literature [45, 47, 53, 54].

A positive stance towards mathematics and physics among early career complexity researchers could be due to the fact that many real-world models of complexity science are grounded in quantitative disciplines such as mathematics and physics [53, 55]. However, a large proportion of complex systems scientists do not identify as mathematicians or physicists, and come from a diverse range of disciplines, including biology, economics, chemistry, archaeology, art, psychology, and the social sciences. The application of quantitative tools to aid the understanding of complex systems might have led to a positive perception of these concepts among professionals, as reflected in their forma mentis network.

Complexity science seems to be a natural candidate for improving the perception of STEM subjects among students. Indeed, previous attempts at building network science courses at the high school level are generally met with interest by high school students [47, 54, 56]. In addition, the Complex Forma Mentis project (www.complexmentis.com Last Accessed: 19 February 2019) provided seminars about complexity science at the high school level that were met with strong interest. Although these initiatives are still early in their implementation, the framework of complexity science may prove to be useful in helping students learn about how the technical aspects of STEM disciplines can be used to address important societal problems, and as a result improve their perception of technical, sometimes obscure, concepts related to mathematical theory and physics.

Another reason behind the dissonant stance of students towards physics and mathematics might be related to a lack of creativity. Recently, Valenti and colleagues [3] measured the implicit attitudes towards science in a population of students and found that the increase of scientific rigour is accompanied by a decrease in associating science with creativity. In the FMN of researchers, "art" is connected to "creative" and "science", whereas these concepts were disconnected in the forma mentis network of students. Researchers also associated "physics" with "creative", an association that is missing in students' FMN. These missing links further underline the importance for students to build a more complete and broader perception of STEM subjects, focusing on the creative process behind science and its real-world, complex implications. Creating such links in high school students, even through simple actions like complexity-focused outreach events, represents a practical task outlined by our results of utmost importance for improving STEM perception.

Nevertheless, the current approach of FMNs has some limitations that we discuss below. The most prominent one is that forma mentis networks operate at the population level, as is common in psycholinguistic approaches relying on free association networks [23, 48] or in educational studies using concept maps [43–45]. Hence, the above patterns have to be interpreted in terms of average trends, as individual students might differ from the aggregated pattern. However, recent approaches have constructed association networks at the level of individuals [31, 36] and even reported how individualized free association networks were predictive of creativity levels [31] or knowledge mastery [36]. With larger sample sizes and a more substantive free association task (leading to denser networks), building forma mentis networks for the individuals represents an exciting research direction for the future.

Another limitation is the experimental effort in engaging participants within a cognitive task, compared to the relative ease of mining online data from social media in order to infer stance. A possible solution could be the use of social media mining to extract semantic associations for forma mentis networks, analogous to the semantic networks of concept co-occurrences in Twitter by [12]. Although this might decrease the difficulty of building a network representation of the mental lexicon of a given population, co-occurrences of words in text are different from free associations and provide different cognitive information with regards to language acquisition and use. For instance, in [20], free associations proved to be more predictive of early word learning compared to word co-occurrences in child directed speech. Hence, despite the difficulty of collection free associations, we argue that free associations provide important insights into the structure of the mental lexicon and that such data are worth the time and cost of data collection.

Potential impact of forma mentis networks in education and beyond

Forma mentis networks represent a powerful new framework that can help tackle important research question within Education research and beyond.

We envision that the most useful educational utilisation of FMNs lies in learning assessment and data-driven educational policy making. The cognitive representation of students' mindsets provided by FMNs represents a powerful way of testing the impact and effectiveness of different teaching methods. Comparisons between the FMNs of a class of students before and after attending a course could provide global and microscopic quantitative information about how students changed their perception and deep understanding of course topics. Furthermore, individual FMNs could be built and correlated with course grades in order to assess the most beneficial changes in mindsets correlating with best exam performances. Promising language network applications of this type have been recently suggested [36, 37, 46] and they confirm the power of network-based representations of knowledge for performance assessment beyond standard tests or quizzes. Once corroborated against individual-level learning performances, FMNs would provide a data-informed approach for facilitating and accelerating conceptual learning based on learners' mindsets.

Forma mentis networks constitute a novel representation of conceptual knowledge and as such can be of great relevance for the understanding of cognition and information processing beyond educational setting. Recently, networks of conceptual knowledge have proved valuable models for understanding knowledge building and exploration in relation to personality traits such as curiosity, openness to experience and creativity [33, 57]. Modelling knowledge acquisition through the statistical mechanics of network walks, de Arruda and colleagues [57] showed that on artificial networks of conceptual associations, knowledge building is consistently stronger in central network regions, i.e. for tightly connected concepts connected by a few steps to all other words. Testing the same dynamics over a "real" mindset represented by a FMN would further characterize the relevance and meaning of key concepts in a given forma mentis. For instance, how relevant "robotics" and "health" can be in healthcare knowledge building across different groups of medical professionals?

Notice that knowledge creation has been recently shown to be driven not only by conceptual centrality/relevance but also by individual personality aspects, such as curiosity (i.e., the proclivity to search for information). It would be interesting to investigate whether the structure of individual FMNs correlate with longitudinal data about curiosity levels. This would allow us to identify distinctive features in mindset organisation around specific topics in high and low curiosity people. Building on the powerful approach by Lydon-Staley and colleagues [34], who recently found that higher curiosity corresponds to tighter networks of associations, FMNs would offer the possibility to assess the impact that positive/negative/neutral concepts and emotional auras play in knowledge building across various levels of curiosity. Another personality trait investigated through associative network approaches was Openness to Experience, the enjoyment of novel ideas and experiences. Christensen and colleagues [35] showed that groups of individuals with a higher Openness to Experience gave rise to a more interconnected network of conceptual associations, thus opening new challenges for characterizing such personality trait in terms of network data. Formulating a predictor of Openness to Experience relying on the conceptual knowledge and sentiment patterns encapsulated in a FMN would complement the descriptive power of forma mentis networks in outlining the stance, and acceptance, a given group has toward a given topic. Another interesting interplay to investigate with FMNs would be the one between knowledge structure and creativity levels. Several recent network approaches managed to correlate the structure of semantic memory with creativity levels [30, 31]. At a population level, Stella and Kenett [32] showed that the multiplex combination of free associations with other semantic, categorical and phonological word-word relationships identifies a central region in the network of conceptual knowledge. The authors found that high/low creativity level people accessed this central region in several different ways. The authors exploited these differences for implementing a machine learning predictor of creativity levels. Replacing the layer of free associations with a forma mentis network and following the protocol by [32] would enable novel ways of testing how creativity levels impact knowledge exploration for different, specific mindsets. Also, building upon our above findings about negative emotional auras correlating with anxiety eliciting and anxiety being a distinctive trait of creative people [8], this application could shed more light on the interplay between the emotional homophily detected in this work, negative/positive sentiment and knowledge exploration across high and low creativity levels.

Conclusions

This article introduced the new methodology of forma mentis networks and demonstrated its potential to identify contrasting stances in different populations. A forma mentis network consists of words as nodes, each with a valence attribute, and free associations as links. Rather than being based on automatic natural language processing, these networks directly access the mental lexicon of human participants, addressing the orthogonal influences of semantic knowledge and emotional affect that drive the processing of information [15, 17, 18] and its consequences [33, 35, 41, 42].

We found substantial differences in the stances of young researchers in complexity science and high school students towards STEM concepts such as physics and mathematics. Students tended to surround these concepts with a negative emotional aura, which could related to a perceived anxiety toward these subjects (cf. [4, 38, 39]). This negative emotional aura was absent in the forma mentis of researchers. Furthermore, the words with a negative valence in the students' neighborhoods were mostly that of mathematical tools, such as "integral" or "function". This result highlighted how the students' negative stance toward physics and mathematics might predominantly originate from an arid view of the tools and methods used in mathematics and physics, which students (but not researchers) perceived as deprived of more interdisciplinary and creative connections.

This quantitative evidence opens new avenues for intervention in education: Encouraging students to incorporate mathematics and physics into a richer association network and drawing new connections to other concepts in the scientific realm represent promising pathways to change their stance. In that lies the potential of the forma mentis network approach–by providing a map of the students' mental lexicon, it is able to show which and where new meaningful links, i.e. associations, could be constructed to maximize the effectiveness of future intervention policies and outreach programmes.

Supporting information

S1 Appendix. Supporting text and figures.
(PDF)

Acknowledgments

The authors acknowledge the Winter Workshop on Complex Systems, the NetSciEd community, the Mediterranean School on Complex Networks and the schools that enthusiastically joined the project: IISS "E. Medi", Galatone (principal: Davide Cammarota), Liceo "G. Banzi Bazoli", Lecce (principal: Antonella Manca), IISS "E. Fermi", Lecce (principal: Giuseppe Russo). Many thanks also to Antonella Conte, Mason A. Porter and Hiroki Sayama for their kind feedback on the manuscript.

Author Contributions

Conceptualization: Massimo Stella, Cynthia S. Q. Siew.

Data curation: Massimo Stella.

Formal analysis: Massimo Stella, Sarah de Nigris, Aleksandra Aloric, Cynthia S. Q. Siew.

Investigation: Massimo Stella, Sarah de Nigris, Aleksandra Aloric, Cynthia S. Q. Siew.

Methodology: Massimo Stella, Sarah de Nigris, Cynthia S. Q. Siew.

Project administration: Massimo Stella, Sarah de Nigris.

Supervision: Massimo Stella.

Validation: Massimo Stella, Aleksandra Aloric, Cynthia S. Q. Siew.

Visualization: Massimo Stella.

Writing – original draft: Massimo Stella, Sarah de Nigris, Aleksandra Aloric, Cynthia S. Q. Siew.

Writing – review & editing: Massimo Stella, Sarah de Nigris, Aleksandra Aloric, Cynthia S. Q. Siew.

References

1. Osborne J, Simon S, Collins S. Attitudes towards science: A review of the literature and its implications. International journal of science education. 2003; 25(9):1049–1079. https://doi.org/10.1080/0950069032000032199
2. Krapp A, Prenzel M. Research on interest in science: Theories, methods, and findings. International journal of science education. 2011; 33(1):27–50. https://doi.org/10.1080/09500693.2010.518645
3. Valenti S, Masnick A, Cox B, Osman C. Adolescents' and Emerging Adults' Implicit Attitudes about STEM Careers:" Science Is Not Creative". Science Education International. 2016; 27(1):40–58.

4. Ashcraft MH. Math anxiety: Personal, educational, and cognitive consequences. Current directions in psychological science. 2002; 11(5):181–185. https://doi.org/10.1111/1467-8721.00196

5. Rothwell J. The hidden STEM economy. Brookings; 2013.

6. Marginson S, Tytler R, Freeman B, Roberts K. STEM: country comparisons: international comparisons of science, technology, engineering and mathematics (STEM) education. Final report. 2013.

7. Aitchison J. Words in the mind: An introduction to the mental lexicon. John Wiley & Sons; 2012.

8. Siegel DJ. The developing mind: How relationships and the brain interact to shape who we are. Guilford Publications; 2015.

9. Gray B, Biber D. Current conceptions of stance. In: Stance and voice in written academic genres. Springer; 2012. p. 15–33.

10. Mohammad S, Kiritchenko S, Sobhani P, Zhu X, Cherry C. Semeval-2016 task 6: Detecting stance in tweets. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016); 2016. p. 31–41.

11. Biber D, Finegan E. Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. Text-interdisciplinary journal for the study of discourse. 1989; 9(1):93–124. https://doi.org/10.1515/text.1.1989.9.1.93

12. Stella M, Ferrara E, De Domenico M. Bots increase exposure to negative and inflammatory content in online social systems. Proceedings of the National Academy of Sciences. 2018; 115(49):12435–12440. https://doi.org/10.1073/pnas.1803470115

13. Somasundaran S, Wiebe J. Recognizing stances in ideological on-line debates. In: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. Association for Computational Linguistics; 2010. p. 116–124.

14. Mohammad SM, Sobhani P, Kiritchenko S. Stance and sentiment in tweets. ACM Transactions on Internet Technology (TOIT). 2017; 17(3):26. https://doi.org/10.1145/3003433

15. Siew CSQ, Wulff DU, Beckage N, Kenett Y. Cognitive Network Science: A review of research on cognition through the lens of network representations, processes, and dynamics. 2018.

16. Steyvers M, Tenenbaum JB. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. Cognitive science. 2005; 29(1):41–78. https://doi.org/10.1207/s15516709cog2901_3 PMID: 21702767

17. Vitevitch MS. What can graph theory tell us about word learning and lexical retrieval? Journal of Speech, Language, and Hearing Research. 2008. https://doi.org/10.1044/1092-4388(2008/030) PMID: 18367686

18. Vitevitch MS, Siew CSQ, Castro N. Spoken Word Recognition. The Oxford Handbook of Psycholinguistics. 2018; p. 31.

19. Stella M, Beckage NM, Brede M, De Domenico M. Multiplex model of mental lexicon reveals explosive learning in humans. Scientific reports. 2018; 8(1):2259. https://doi.org/10.1038/s41598-018-20730-5 PMID: 29396497

20. Stella M, Beckage NM, Brede M. Multiplex lexical networks reveal patterns in early word acquisition in children. Scientific reports. 2017; 7:46730. https://doi.org/10.1038/srep46730 PMID: 28436476

21. Hills TT, Siew CSQ. Filling gaps in early word learning. Nature Human Behaviour. 2018; 2(9):622. https://doi.org/10.1038/s41562-018-0428-y PMID: 31346280

22. Stella M. Modelling Early Word Acquisition through Multiplex Lexical Networks and Machine Learning. Big Data and Cognitive Computing. 2019; 3(1):10. https://doi.org/10.3390/bdcc3010010

23. De Deyne S, Navarro DJ, Storms G. Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. Behavior research methods. 2013; 45(2):480–498. https://doi.org/10.3758/s13428-012-0260-7 PMID: 23055165

24. Kenett YN, Levi E, Anaki D, Faust M. The semantic distance task: Quantifying semantic distance with semantic network path length. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2017; 43(9):1470. https://doi.org/10.1037/xlm0000391 PMID: 28240936

25. De Deyne S, Navarro DJ, Perfors A, Brysbaert M, Storms G. The "Small World of Words" English word association norms for over 12,000 cue words. Behavior research methods. 2018; p. 1–20.

26. Amancio DR. A complex network approach to stylometry. PloS one. 2015; 10(8):e0136076. https://doi.org/10.1371/journal.pone.0136076 PMID: 26313921

27. Akimushkin C, Amancio DR, Oliveira ON Jr. Text authorship identified using the dynamics of word co-occurrence networks. PloS one. 2017; 12(1):e0170527. https://doi.org/10.1371/journal.pone.0170527 PMID: 28125703

28. Zemla JC, Austerweil JL. Analyzing Knowledge Retrieval Impairments Associated with Alzheimer's Disease Using Network Analyses. Complexity. 2019; 2019. https://doi.org/10.1155/2019/4203158 PMID: 31341377

29. Castro N, Stella M. The multiplex structure of the mental lexicon influences picture naming in people with aphasia. Journal of Complex Networks. 2019;. https://doi.org/10.1093/comnet/cnz012

30. Kenett YN, Anaki D, Faust M. Investigating the structure of semantic networks in low and high creative persons. Frontiers in human neuroscience. 2014; 8:407. https://doi.org/10.3389/fnhum.2014.00407 PMID: 24959129

31. Kenett YN, Levy O, Kenett DY, Stanley HE, Faust M, Havlin S. Flexibility of thought in high creative individuals represented by percolation analysis. Proceedings of the National Academy of Sciences. 2018; 115(5):867–872. https://doi.org/10.1073/pnas.1717362115

32. Stella M, Kenett YN. Viability in Multiplex Lexical Networks and Machine Learning Characterizes Human Creativity. Big Data and Cognitive Computing. 2019; 3(3):45. https://doi.org/10.3390/bdcc3030045

33. Zurn P, Bassett DS. On Curiosity: A Fundamental Aspect of Personality, a Practice of Network Growth. Personality Neuroscience. 2018; 1. https://doi.org/10.1017/pen.2018.3

34. Lydon-Staley DM, Zhou D, Blevins AS, Zurn P, Bassett DS. Hunters, busybodies, and the knowledge network building associated with curiosity.

35. Christensen AP, Kenett YN, Cotter KN, Beaty RE, Silvia PJ. Remotely close associations: Openness to experience and semantic memory structure. European Journal of Personality. 2018; 32(4):480–492. https://doi.org/10.1002/per.2157

36. Siew CSQ. Using network science to analyze concept maps of psychology undergraduates. Applied Cognitive Psychology. 2018;. https://doi.org/10.1002/acp.3484

37. Valenzuela Castellanos MF, Pérez Villalobos M, Bustos C, Salcedo Lagos P. Cambios en el concepto aprendizaje de estudiantes de pedagogía: análisis de disponibilidad léxica y grafos. Estudios filológicos. 2018;(61):143–173.

38. Siew CSQ, McCartney MJ, Vitevitch MS. Using network science to understand statistics anxiety among college students. Scholarship of Teaching and Learning in Psychology. 2019;. https://doi.org/10.1037/stl0000133

39. Posner J, Russell JA, Peterson BS. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and psychopathology. 2005; 17(3):715–734. https://doi.org/10.1017/S0954579405050340 PMID: 16262989

40. Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13,915 English lemmas. Behavior research methods. 2013; 45(4):1191–1207. https://doi.org/10.3758/s13428-012-0314-x PMID: 23404613

41. Adelman JS, Estes Z. Emotion and memory: A recognition advantage for positive and negative words independent of arousal. Cognition. 2013; 129(3):530–535. https://doi.org/10.1016/j.cognition.2013.08.014 PMID: 24041838

42. Gaillard R, Del Cul A, Naccache L, Vinckier F, Cohen L, Dehaene S. Nonconscious semantic processing of emotional words modulates conscious access. Proceedings of the National Academy of Sciences. 2006; 103(19):7524–7529. https://doi.org/10.1073/pnas.0600584103

43. Koponen IT, Pehkonen M. Coherent knowledge structures of physics represented as concept networks in teacher education. Science & Education. 2010; 19(3):259–282. https://doi.org/10.1007/s11191-009-9200-z

44. Koponen IT, Nousiainen M. Concept networks of students' knowledge of relationships between physics concepts: finding key concepts and their epistemic support. Applied Network Science. 2018; 3(1):14. https://doi.org/10.1007/s41109-018-0072-5

45. Sayama H, Cramer C, Porter MA, Sheetz L, Uzzo S. What are essential concepts about networks? Journal of Complex Networks. 2016; 4(3):457–474. https://doi.org/10.1093/comnet/cnv028

46. Koponen IT, Nousiainen M. Pre-Service Teachers' Knowledge of Relational Structure of Physics Concepts: Finding Key Concepts of Electricity and Magnetism. Education Sciences. 2019;. https://doi.org/10.3390/educsci9010018

47. Cramer CB, Porter MA, Sayama H, Sheetz L, Uzzo SM. Network Science In Education: Transformational Approaches in Teaching and Learning. 1st ed. Springer Publishing Company, Incorporated; 2018.

48. Nelson DL, McEvoy CL, Schreiber TA. The University of South Florida free association, rhyme, and word fragment norms. Behavior Research Methods, Instruments, & Computers. 2004; 36(3):402–407. https://doi.org/10.3758/BF03195588

49. Fairfield B, Ambrosini E, Mammarella N, Montefinese M. Affective norms for Italian words in older adults: age differences in ratings of valence, arousal and dominance. PloS one. 2017; 12(1):e0169472. https://doi.org/10.1371/journal.pone.0169472 PMID: 28046070

50. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: Homophily in social networks. Annual review of sociology. 2001; 27(1):415–444. https://doi.org/10.1146/annurev.soc.27.1.415

51. Laukenmann M, Bleicher M, Fuß S, Gläser-Zikuda M, Mayring P, von Rhöneck C. An investigation of the influence of emotional factors on learning in physics instruction. International Journal of Science Education. 2003; 25(4):489–507. https://doi.org/10.1080/09500690210163233

52. Lehtamo S, Juuti K, Inkinen J, Lavonen J. Connection between academic emotions in situ and retention in the physics track: applying experience sampling method. International journal of STEM education. 2018; 5(1):25. https://doi.org/10.1186/s40594-018-0126-3 PMID: 30631715

53. Resnick M. Turtles, termites, and traffic jams: Explorations in massively parallel microworlds. Mit Press; 1997.

54. van der Cingel P. How to educate navigators in a complex world: making a case in higher professional education in the Netherlands. Complexity, governance and networks. 2018; 4(1).

55. Mitchell M. Complexity: A guided tour. Oxford University Press; 2009.

56. Cramer C, Gera R, Panagakou E, Porter MA, Sayama H, Sheetz L, et al. Proceedings of NetSciEd 2018. OSF Preprints; 2018.

57. de Arruda HF, Silva FN, Costa LdF, Amancio DR. Knowledge acquisition: A Complex networks approach. Information Sciences. 2017; 421:154–166. https://doi.org/10.1016/j.ins.2017.08.091

ROYAL SOCIETY
OPEN SCIENCE

Research

Cite this article: Nicole R, Alorić A, Sollich P. 2021 Fragmentation in trader preferences among multiple markets: market coexistence versus single market dominance. *R. Soc. Open Sci.* **8**: 202233.
https://doi.org/10.1098/rsos.202233

Received: 7 January 2021
Accepted: 2 August 2021

Subject Category:
Computer science and artificial intelligence

Subject Areas:
statistical physics/computer modelling and simulation/complexity

Keywords:
agent-based modelling, reinforcement learning, double auction markets, large deviations, effective single agent dynamics, Fokker–Planck equation

Author for correspondence:
Aleksandra Alorić
e-mail: aleksandra.aloric@gmail.com

[1]Department of Mathematics, King's College London, Strand, London WC2R 2LS, UK
[2]Scientific Computing Laboratory, Center for the Study of Complex Systems, Institute of Physics Belgrade, University of Belgrade, Pregrevica, 118, 11080 Belgrade, Serbia
[3]Institut für Theoretische Physik, Georg-August-Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

AA, 0000-0002-7278-599X

Technological advancement has led to an increase in the number and type of trading venues and a diversification of goods traded. These changes have re-emphasized the importance of understanding the effects of market competition: does proliferation of trading venues and increased competition lead to dominance of a single market or coexistence of multiple markets? In this paper, we address these questions in a stylized model of zero-intelligence traders who make repeated decisions at which of three available markets to trade. We analyse the model numerically and analytically and find that the traders' decision parameters—memory length and how strongly decisions are based on past success—make the key difference between consolidated and fragmented steady states of the population of traders. All three markets coexist with equal shares of traders only when either learning is too weak and traders choose randomly, or when markets are identical. In the latter case, the population of traders fragments across the markets. With different markets, we note that market dominance is the more typical scenario. Overall we show that, contrary to previous research emphasizing the role of traders' heterogeneity, market coexistence can emerge simply as a consequence of co-adaptation of an initially homogeneous population of traders.

THE ROYAL SOCIETY
PUBLISHING

The possible risks and benefits of market competition have been the subject of a long-standing debate, which is often expressed as 'market consolidation versus market fragmentation' [1,2]. When the New York Stock Exchange had by far the strongest influence on price formation, the financial trading system was much closer to a consolidated state (Hasbrouck's [3]), but more recently technological progress has created a variety of trading venues and led to ever-increasing market fragmentation. Particularly interesting in this regard are the so-called *dark pools*. These trading venues have gained a certain notoriety from their lack of transparency and the possibility to trade large volumes without large price impacts, and they frequently offer a greater variety of market mechanisms compared to the conventional exchanges. Shorter & Miller [4] noted that in only five years (from 2008 to 2013) the US market share traded in dark pools increased from 4% to 15%, signalling a distinct increase in market fragmentation. Gomber *et al.* [1] suggest that the main driver of market fragmentation is the heterogeneity of traders' needs, which will be more easily satisfied by a variety of different markets rather than a single trading venue. In this paper, we show that even when identical markets compete, economic agents can develop loyalties to specific markets, thus effectively fragmenting trading. Conversely, we find in the case of competition of markets that are biased towards different classes within the population of traders, single market dominance is the typical outcome.

To tackle this question of market coexistence versus single market dominance, we build on previous work [5–8] where we introduced and analysed a system consisting of double auction markets and a large number of traders choosing between them. What we showed in this setting is that for a range of parameters describing the markets and agents, the agents split into groups with a strong loyalty towards one of the markets, often giving an overall market coexistence with an equal share of traders at both markets. When the agents have a long memory to previous trading outcomes, other steady states with single market dominance also exist and are in fact stable, whereas the system state with markets splitting trades roughly equally between them is only metastable [6,8]. While these initial studies focused on settings with two markets for simplicity, traders do in general have a choice between multiple markets (e.g. [1]) and this feature was also present in the CAT game [9] that originally motivated our research into market-trader co-fragmentation. We therefore extend the double auction market model from two to three markets in this paper, and use the results to formulate conjectures for the expected behaviour in cases where more than three markets compete.

There is a large body of work that uses the *JCAT* library [10] to explore competition between *continuous double auction markets* [11–13]. In a spirit similar to our work, they use simple learning algorithms such as Zero-Intelligence [14] or Zero-Intelligence-Plus [15] for both markets and traders, and analyse the allocation efficiency of double auction markets when they are competing against each other. Multi-agent-based simulations have mostly been used in this context and allow additional layers of complexity such as *adaptive markets* and *heterogeneous agents* to be added. We pursue instead a modelling approach that strips out as much detail as possible [6–8] to allow for detailed theoretical analysis, which can often reveal features that would be missed when relying exclusively on numerical simulations. In this spirit, while the market mechanisms implemented in the JCAT library are *continuous* double auctions, we use in our model a mechanism more similar to a *clearing house* where the clearing process takes place at discrete time steps. This makes a largely analytical approach possible, which reveals the learning process of the agents as the main driver of fragmentation. This conclusion was shown in [6] to carry over to models with more complex market mechanisms and more sophisticated agent strategies, based e.g. on [16].

Authors such as Ellison *et al.* [17] and Shi *et al.* [18] have focused on studying the competition between markets and the conditions under which this led to multiple market coexistence or the emergence of a market monopoly. The authors name two significant effects in the competition of double auctions, one of them is the positive size effect, i.e. agents prefer trading in a market where there are already many traders of the opposite type (e.g. sellers like trading at markets where there are many buyers), as the choice among offers is better. The authors additionally suggest the existence of a negative size effect in a double auction market, as agents will prefer being in the minority group to trade more often (e.g. buyers see the benefit of trading at a market where there are not many buyers, e.g. [19]). Ellison *et al.* [17] point out that due to this negative size effect, coexistence of many markets is possible. On the other hand, Shi *et al.* [18] investigate which of the two effects is stronger and finds that due to more substantial positive effects, a monopoly will in many situations be the preferred outcome. When there is strong market differentiation, Shi *et al.* [18] argue that market coexistence is possible, especially for markets that have different pricing policies, e.g. where one market charges a fixed participation fee while another charges a profit fee. Although in what follows we will consider markets without fee charging policies, we will find nonetheless there are system parameter ranges that enable coexistence, where markets are populated by roughly the same numbers of traders; conversely, we also identify the parameter regimes for which one market is dominant. It is important to note that the studies cited above have focused on finding either the Nash equilibria or states favoured by the replicator dynamics. By contrast, we consider dynamics based on agents learning to improve their market choosing strategy, which we believe is more appropriate in the context of agents engaging in economic interactions. In this study, we show that fragmentation can arise even in an initially homogeneous population of traders, only because the traders adapt to their past record of successful trades.

Here we summarize the basic assumptions and properties of the model introduced in [5,6,8] and extend it to include multiple markets.

We study a population of agents without sophisticated trading strategies, essentially zero-intelligence traders [14,20,21]. The orders to buy at a certain price (bids) and orders to sell at a certain price (asks) are assumed to be unrelated to previous trading success or any other information. We assume that bids, b, and asks, a, are normally distributed ($a \sim \mathcal{N}(\mu_a, \sigma_a^2)$ and $b \sim \mathcal{N}(\mu_b, \sigma_b^2)$), where $\mu_b > \mu_a$, in line with [6]. After each round of trading each agent receives a score, reflecting their payoff in the trade. The scores of agents who do trade are assigned as elsewhere in the literature [14,22]: buyers value paying less than they offered (b), and so their score is $S = b - \pi$, where π is the trading price. Sellers value trading for more than their ask (a), and so $S = \pi - a$ is a reasonable model for their payoff.

The role of a market is to facilitate trades so we define markets in terms of their price-setting and order-matching mechanisms. We consider a single-unit discrete time double auction market where all orders arrive simultaneously and market clearing happens once every period after the orders are collected. We also assume that a uniform price is set by the market—once all orders have arrived, these are used to determine average bid $\langle b \rangle$ and average ask $\langle a \rangle$ and then set a global trading price in between the two

$$\pi = \langle a \rangle + \theta(\langle b \rangle - \langle a \rangle), \tag{2.1}$$

where θ fixes the price closer to the average bid ($\theta > 0.5$) or the average ask ($\theta < 0.5$); the parameter θ thus represents the bias of the market towards sellers (they earn more when $\theta > 0.5$) or buyers (earn more when $\theta < 0.5$).[1] Once the trading price has been set, all bids below this price, and all asks above it, are marked as invalid orders that cannot be executed at the current trading price. The remaining orders are executed by randomly pairing buyers and sellers; the execution price is π. Note that we assume here that each order is for a single unit of the good traded.

The most efficient resource allocation happens when demand equals supply, i.e. at the equilibrium trading price. In a set-up like ours where the bids and asks are Gaussian random variables with equal variances ($\sigma_a = \sigma_b$) and when the number of buyers is equal to the number of sellers at a given market, the equilibrium trading price corresponds to $\theta = 0.5$, i.e. the price is $\pi^{\text{eq}} = (\langle b \rangle + \langle a \rangle)/2$. We start off below by considering such efficient markets and will also call these *fair* as $\theta = 0.5$; later we allow for the possibility that markets are not fair and set the price closer to the average bid or ask ($\theta \neq 0.5$).

Agents trade repeatedly in our model, and they adapt their preferences for the various choices at their disposal from one trading period to the next. We assume that each agent decides where to trade

[1] Note that traders are not informed about these market biases, nor the market mechanism in general; they only obtain information through the scores they receive.

(which of many markets) at the beginning of each trading period, only based on his or her past experience. To formalize this we introduce a set of attractions A_m for each player, one for each market $m = 1, 2, 3$. The attractions will generally differ from player to player, but we suppress this in the notation for now. The attractions are updated after every trading period, n, using the following reinforcement learning rule (similar to Q-learning [23] and the experience-weighted attraction rule [24,25])

$$A_m(n+1) = \begin{cases} (1-r)A_m(n) + rS_m(n) & \text{if the agent chose market } m \text{ in round } n \\ (1-r)A_m(n) & \text{otherwise.} \end{cases} \tag{2.2}$$

The quantity $S_m(n)$ is the score gained trading at market m in the nth trading period. The length of the agents' memory is set by r: effectively an agent takes into account a sliding window of length of order $1/r$ for the weighted averaging of past returns.

Once each preference is updated, traders use the *multinomial logit function* to choose at which market to trade in the next round

$$P(M = m) = \frac{\exp(\beta A_m)}{\sum_{m'} \exp(\beta A_{m'})}. \tag{2.3}$$

This is inspired by the experience-weighted attraction literature [24,25], where β is the *intensity of choice* and regulates how strongly the agents bias their preferences towards actions with high attractions. For $\beta \to \infty$, the agents choose the option with the highest attraction, while for $\beta \to 0$ they choose randomly with equal probabilities among all options.

Agents randomly take the role of buyer or seller in each trading round: they act as buyers with probability $p_{\mathcal{B}}$, which we call their buying preference. We will study a population of traders consisting of two classes of agents with fixed buying preferences $p_{\mathcal{B}} = p_{\mathcal{B}}^{(1)}$ and $p_{\mathcal{B}} = p_{\mathcal{B}}^{(2)}$, respectively. The attractions of agents from different classes will be denoted by $A_m^{(c)}$ with $c \in \{1, 2\}$.

We will frequently study a set-up with symmetric markets (i.e. $\theta_1 = 1 - \theta_2 < 0.5$) and a population consisting of two symmetrically biased classes (i.e. $p_{\mathcal{B}}^{(1)} = 1 - p_{\mathcal{B}}^{(2)} > 0.5$). The setting considered as default in [6] is $(\theta_1, \theta_2, p_{\mathcal{B}}^{(1)}, p_{\mathcal{B}}^{(2)}) = (0.3, 0.7, 0.8, 0.2)$. It is such that the class 1 (buyers) prefer trading at market 1, that is biased to award buyers with higher returns, while agents of class 2 (sellers) prefer market 2. It has been shown previously that for low intensity of choice β, the unique fixed point of the learning dynamics is such that agents develop a higher attraction to the market that is better for them; nonetheless, they trade largely at random because of the low β. When β is increased, this fixed point becomes unstable as buyers and sellers would congregate in different markets and so lose many trading opportunities. Instead the population fragments: agents of both classes self-organize to divide into two groups within each class. One of these groups is return oriented (e.g. buyers at market 1) and the corresponding agents earn more per single trade; the other group can be characterized as volume oriented (e.g. sellers at market 1), earning less per trade but having the opportunity to trade more often.

To motivate the use of this stylized model of agents choosing between multiple markets, we start with multi-agent simulations of the system. We look at a default population of traders consisting of two classes—some tend to act more as buyers ($p_{\mathcal{B}} = 0.8$), others more as sellers ($p_{\mathcal{B}} = 0.2$). These traders choose between three markets that differ in their biases θ. We show an example of three qualitatively different distributions of the attractions of the agents in figure 1. To facilitate the interpretation of these distributions, we mark by coloured regions in each panel which market an agent prefers at the given attraction (differences), i.e. which market s/he chooses with the highest probability.

We now give a brief description of the attractions distributions in each of the panels and explain the difference between (i) strong fragmentation, which persists in the large memory limit, and (ii) weak fragmentation, which disappears in the same limit; similar results for two market systems are discussed in [6,8]. In figure 1*a*, one sees that the distribution of attractions has three peaks, all of which have a size of order $O(1)$ and correspond to subpopulations of traders who choose to trade mainly at a single market. In other words, the trader population (in the class shown in the figure) splits into three subpopulations that are more attracted to one market over the others, e.g. traders develop individual loyalties to one of the markets. Such distributions of attractions with more than one peak with a size of order one are called *strongly fragmented* [8]. As discussed in previous works, this does not mean the traders' preferences are frozen: they do change their preferred market but only

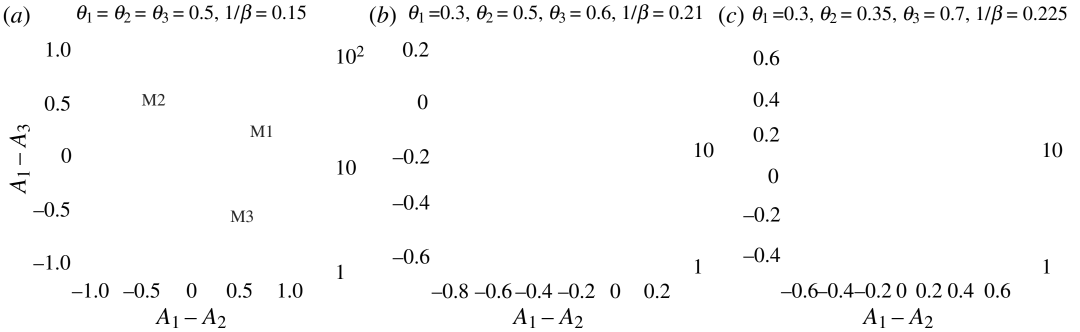

Figure 1. Distribution of attraction differences of population of traders for market and learning parameters as indicated in each graph title. In (*a*), the population is strongly fragmented into three groups of equal size. In (*b*), the population is weakly fragmented, the distribution has two peaks: one large peak and one peak that (as we will later see) becomes exponentially small as the memory length increases. In (*c*), the population is strongly fragmented, but only across two markets. To obtain those graphs, we ran simulations with $r = 0.01$ and $N/2 = 10\,000$ traders in each class until a steady state was reached. Traders from class 1 have preference to buy $p_{\mathcal{B}}^{(1)} = 0.8$ and traders from class 2 have preference to buy $p_{\mathcal{B}}^{(2)} = 0.2$. The $(A_1 - A_2, A_1 - A_3)$ plane is shown subdivided into three zones that indicate which market an agent with the corresponding attractions chooses most often. The zones are coloured blue, red and green for markets 1, 2 and 3, respectively, as indicated in (*a*).

after a long persistence time [6]. We also note that in the state shown, i.e. for the given parameters, three identical markets coexist and receive an equal share of traders, on average.

The second distribution, shown in panel (*b*), corresponds to a population divided into two loyalty groups but with different sizes: one large (order N) subpopulation is attracted to the second market (the fair market, $\theta = 0.5$), while the second, smaller subpopulation persistently tries to trade at market 3. The size of the smaller peak in the attraction distribution decreases exponentially as $r \to 0$ [7,8], and although markets 2 and 3 coexist for any finite r, in the large memory limit, market 2 has a monopoly. When attraction distributions are multimodal but only one peak has a weight of order 1 (i.e. fragmentation is only present at finite r) we call them *weakly fragmented*.

The distribution plotted in panel (*c*) corresponds to a strongly fragmented population, but contrary to the case depicted in panel (*a*) the third market has now lost the competition. Additionally, the share of attracted traders is not the same between the markets (as in panel (*a*)), but both peaks persist in the long memory limit.

The above simulation results offer a glimpse into a rich variety of qualitatively different structures of the attraction distributions (number and size of peaks) and consequently different outcomes of a three-market competition. To study these in more detail, we focus on the analytical and numerical methods described previously [7,8] for large populations of traders and in the large memory limit ($r \to 0$).

To proceed with the analysis, in line with our earlier studies [6–8], we start from the fact that the system is Markovian and accordingly the master equation introduced in [6] is an exact and complete description of the evolution of agents in the limit of an infinite population N and large memory $1/r$. We focus here on the steady states of this dynamical evolution. For a population with fixed buy/sell preferences, this is specified by a steady-state distribution $P(\mathbf{A}|p_{\mathcal{B}})$ where $\mathbf{A}$ is an M-dimensional vector of attractions and conditioning on the buying preference and distinguishes the different classes of traders. When we study more than two markets the distribution is multivariate, though we can introduce attraction differences and look for a solution in the resulting $M - 1$ variables. The master equation describing the evolution of the system [6] across the different trading rounds n is not a standard linear Chapman–Kolmogorov equation as the transition kernel K depends on the trading probabilities, which in turn depend on $P_n(\mathbf{A}|p_{\mathcal{B}})$. This self-consistent nature of the description arises from the reduction from a description in terms of the attractions of all N agents to one for a single agent; this reduction becomes exact for $N \to \infty$. In principle, a steady state could then be found by tracking the evolution in time from the initial condition $P_0(\mathbf{A}|p_{\mathcal{B}}) = \delta(\mathbf{A})$, which corresponds to all agents having zero attraction to all markets. We take a different route and first transform the time evolution equation to a Fokker–Planck description using the Kramers–Moyal expansion. This is appropriate for small r, i.e. for agents with long memory.

Even after the simplification to a Fokker–Planck equation, the dimensionality of the problem makes finding the steady state a non-trivial task. But we can make progress by considering the limit $r \to 0$; this will allow us to evaluate the onset of fragmentation. We do this by analysing the drift $\mu_m^{(c)}$ in the Fokker–Planck equation, defined in appendix A. To find the single agent steady state, we will search for zeros of the drift assuming fixed market order parameters, i.e. trading probabilities. We start by assuming that the two classes have homogeneous preferences for the markets (i.e. $P(\mathbf{A}^{(c)}|p_B^{(c)})$ is a delta distribution). This is the expected solution in the low β-limit, when the steady state is unfragmented. With this assumption, the expressions for the market order parameters simplify, and we can solve the simultaneous equations for the two classes. At any fixed point solution $(\mathbf{A}^{(1)^*}, \mathbf{A}^{(2)^*})$ we evaluate the market order parameters and check if the single agent dynamics is consistent with the homogeneous population assumption: when we solve $\mu_m^{(c)}(\mathbf{A}) = 0$ we expect only one zero that coincides with $(\mathbf{A}^*)$. The onset of fragmentation (weak or strong) is then given by the intensity of choice where the single agent dynamics first has multiple zeros when evaluated at the homogeneous population market order parameters, which indicates that for $r > 0$ the distribution of attractions will have multiple peaks. To find the weights of the attraction distribution at each peak, corresponding to a fixed point, we use the Freidlin–Wentzell approach detailed in appendix B. This allows us to differentiate between small peaks, which decay exponentially with the memory length $1/r$, and large peaks, whose weight remains finite and of order unity when the $r \to 0$ limit is taken.

In the rest of the paper, we focus our analysis on a scenario with $M = 3$ markets and we describe each of the two classes in terms of the two attraction differences $\Delta A_2 = A_1 - A_2$ and $\Delta A_3 = A_1 - A_3$. We perform a Kramers–Moyal expansion of the trader's learning dynamics and obtain two Fokker–Planck equations (one for each class $c \in \{1, 2\}$ of traders) for the distribution of attraction differences $P(\mathbf{\Delta A}^{(c)}, t)$

$$\begin{aligned}\partial_t P(\mathbf{\Delta A}^{(c)}, t) = &-\sum_{m=2}^{3} \partial_{\Delta A_m^{(c)}} [\mu_m^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3) P(\mathbf{\Delta A}^{(c)}, t)] \\ &+ \frac{r}{2} \sum_{m,m'=2}^{3} \partial_{\Delta A_m^{(c)}} \partial_{\Delta A_{m'}^{(c)}} [\Sigma_{mm'}^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3) P(\mathbf{\Delta A}^{(c)}, t)]. \qquad (3.1)\end{aligned}$$

Here the time variable $t = nr$ is a rescaled number of trading rounds, $\mathbf{\Delta A}^{(c)} = (\Delta A_2^{(c)}, \Delta A_3^{(c)})$ and f_m is the market order parameter, i.e. the ratio of buyers to sellers at market m (effectively the demand-to-supply ratio). The expressions for the drift vectors $\mu_m^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3)$ and the covariance matrices $\Sigma_{mm'}^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3)$ for each class are given in appendix A.

We start by looking at what happens when the three markets available are all fair, i.e. $\theta_1 = \theta_2 = \theta_3 = 0.5$. This means they set their trading price to be exactly the mean of the average bid and the average ask. As mentioned previously, the fair market corresponds to a market mechanism delivering the equilibrium trading price, provided the number of buyers equals number of sellers.

Based on intuition from similar physical systems, one might expect spontaneous symmetry breaking, where random fluctuations lead the whole population to select only one of the possible symmetric markets. However, in stochastic multi-agent simulations we observe instead steady states with fragmented populations within each class; we therefore focus on steady states of the traders' learning dynamics without symmetry breaking.

Since the three markets have the same bias θ, in a symmetric solution, they should attract the same number of agents, irrespective of their class. On the other hand, as we study classes of agents with symmetric preferences to buy $p_B^{(1)} = 1 - p_B^{(2)}$, the difference between the number of buyers and the number of sellers at a single market is of order $\sqrt{N}$, $N_B = N_S + O(\sqrt{N})$. As a consequence, in the large size limit, the ratio of the number of buyers to the number of sellers in each market is equal to 1. This simplification is the reason why we choose to start the analysis with the simple case of three fair markets, which allows one to explore the phenomenon of fragmentation across three double auction markets without the need for a self-consistent determination of market order parameters [7,8].

We start by looking at the fixed point structure of the single agent dynamics when the intensity of choice β is small. As expected, the only fixed point of the learning dynamics is $A_1 - A_2 = A_1 - A_3 = 0$ and corresponds to a trader who chooses to randomize between the three markets (figure 2*a*). When the intensity of choice β reaches a critical value $\beta_c = 1/0.254$, three saddle node bifurcations take place simultaneously and three pairs of stable and unstable fixed points appear (figure 2*b*). The reason why

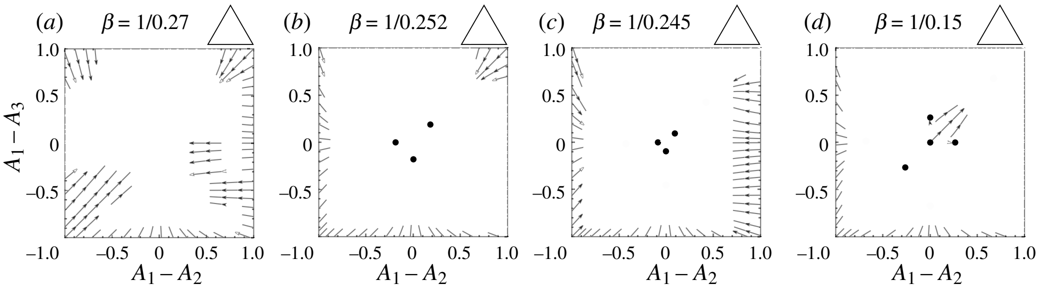

Figure 2. Flow diagram and fixed points of the learning dynamics of a single trader with $p_B^{(2)} = 0.2$, choosing between three fair markets. (*a*) Below the weak fragmentation threshold $\beta = 1/0.254$, the dynamics only has one fixed point, which is stable (denoted with red star). (*b*) When β reaches the weak fragmentation threshold $\beta_c = 1/0.254$, three pairs of unstable (blue) and meta-stable (red empty circles) fixed points appear and the system becomes weakly fragmented with one large peak, which corresponds to traders randomizing between the three markets, and three small peaks where agents trade preferentially at one of the three available markets. (*c*) At $\beta_c' = 1/0.252$, the three outer fixed points become stable and the central one meta-stable and the system is now strongly fragmented, with three peaks of equal size each of which corresponds to preferentially trading at a single market. (*d*) As β increases, the meta-stable fixed points eventually becomes unstable. Above each graph we indicate in triangular notation the category to which each of the fixed point structures belongs (see main text for details).

those three saddle node bifurcations take place at the same time lies in the markets' symmetry, i.e. their identical bias $\theta = 0.5$. In the more general case where the three markets are different, we expect the appearance of each pair of new fixed points to take place at a different value of β.

When looking at the deterministic dynamics for low intensity of choice (figure 2*a*), it is obvious that the system is not fragmented and there is only one stable fixed point. At larger intensities of choice as in figure 2*b–d*, knowing the deterministic dynamics is not sufficient to distinguish between 'stable' fixed points (the ones where, in our terminology, large peaks will be centred) and 'metastable' ones (which for us indicate the positions of the small peaks). To assess the stability of fixed points in figure 2 and weight sizes of potential peaks, we use the Freidlin–Wentzell approach detailed in appendix B.

As an example of an attraction distribution that has both small and large peaks we consider the range $1/0.252 \geq \beta \geq 1/0.254$ for the intensity of choice, where the system is weakly fragmented (as in figure 2*b*). The central fixed point is stable and a large peak in the attraction distribution is located at this fixed point, while the three outer fixed points are metastable and correspond to small peaks. As β is increased to a second critical value of $\beta_c' = 1/0.252$, the three outer fixed points become stable and the system undergoes a strong fragmentation transition. For any values of β above this second fragmentation threshold, the system will be strongly fragmented as the distribution of preferences of the traders will have three peaks of equal weight, each of which corresponds to a stable fixed point of the single agent dynamics (red points in figure 2*c*,*d*). For $1/0.237 \leq \beta \leq 1/0.252$, the distribution of attractions retains an additional peak at the fixed point at (0, 0) but the weight of this peak will become exponentially small as the memory length increases (figure 2*c*). This metastable fixed point and the associated small peak in the attraction distribution then disappear for $\beta \geq \beta_c'' = 1/0.237$ (figure 2*d*).

We summarize briefly the intuitive meaning of the above results for the attraction distributions in a system of agents with long memory choosing between three fair markets. When the intensity of choice is small the agents cannot develop strong attractions to any particular market as low β implies that they choose a market largely randomly. With increasing β, three small subpopulations of the agents in each class develop a loyalty to one of the markets, signalled by increased attractions, but the random choice strategy remains dominant. These loyal subpopulations grow until (beyond β_c') they encompass most of each agent class.

To help with understanding the variety of different steady states, we introduced an attraction distribution notation in the shape of triangles, as depicted in panels of figure 2. We focus on the number and size of the peaks, rather than their exact position, and use the triangle to visualize attraction to any of the three markets (circle close to the corner) or market indifference (star shape). To distinguish between large and small peaks we use filled or empty objects (both stars and circles).

In the simple case of three competing markets considered so far, we find that they always coexist, but in different scenarios ranging from all traders choosing a market randomly to traders splitting into subpopulations with persistent market loyalties. An obvious question is then whether this fragmentation is critically dependent on the fact that all the markets are identical. To answer this, we next extend our analysis to markets with different biases.

Each market bias θ_1, θ_2, θ_3 is between zero and one, i.e. the market parameter space is a unit cube. Of course the phenomenon of fragmentation is independent under permutation of the market biases as this effectively just changes the labelling of the markets. We can therefore restrict our analysis to 1/6 of the cube where $\theta_1 \leq \theta_2 \leq \theta_3$ and can reconstruct the behaviour in the rest of the parameter space by symmetry. We will mostly follow this scheme but sometimes allow a different parameter ordering to get simpler two-dimensional phase diagrams, with a typical bias along the x-axis and the inverse intensity of choice along the y-axis. We study three different types of scenarios, guided by explorations in our previous work: (i) one fair market $\theta_2 = 0.5$ and two symmetrically biased markets $\theta_1 = 1 - \theta_3$, with θ_1 as a free parameter varying between 0 and 1/2, shown in figure 3, (ii) two symmetrically biased markets $\theta_1 = 0.3$, $\theta_2 = 0.7$ with θ_3 varied as a free parameter, shown in figure 4, (iii) $\theta_1 = 0.3$, $\theta_2 = 0.5$ and θ_3 again ranging from 0 to 1, shown in figure 6. As will become clear in the rest of this section, these parameter settings allow for the analysis of the effect of a number of properties on the occurrence of fragmentation, such as the market symmetry, the 'distance' between market biases and the effect of market fairness.

Following the reasoning we used in the case of three fair markets, we continue to focus on solutions that do not break the market symmetries. This assumption is supported by stochastic multi-agent simulations in which we do not observe market symmetry breaking. We use the symmetries to restrict the possible values of the 'market aggregates', i.e. the demand-to-supply ratios. In particular, we can show that these ratios are inverses of each other for the symmetrically biased markets, and that the ratio is unity at the fair market as before. To see this, note first that when $\theta_1 = 1 - \theta_3$ and $\theta_2 = 0.5$, for traders with symmetric preferences to buy, the role played by market 1 for traders from class 1 is the same as the role played by market 3 for traders from class 2 and vice versa. As a consequence, the probability of trading at the first market for a trader from class 1 (resp. 2) is equal to the probability of trading at the third market for a trader of class 2 (resp. 1). We can write the buyer/seller ratios in market 1 and 3 as

$$\left.\begin{aligned} f_1 &= \frac{P^{(1)}(M=1)p_B^{(1)} + P^{(2)}(M=1)p_B^{(2)}}{P^{(1)}(M=1)(1-p_B^{(1)}) + P^{(2)}(M=1)(1-p_B^{(2)})} \\ \text{and} \qquad f_3 &= \frac{P^{(1)}(M=3)p_B^{(1)} + P^{(2)}(M=3)p_B^{(2)}}{P^{(1)}(M=3)(1-p_B^{(1)}) + P^{(2)}(M=3)(1-p_B^{(2)})}. \end{aligned}\right\} \tag{4.1}$$

When substituting into these expressions the equalities $P^{(1)}(M=1) = P^{(2)}(M=3)$, $P^{(2)}(M=1) = P^{(1)}(M=3)$ and remembering that $p_B^{(1)} = 1 - p_B^{(2)}$, one sees that $f_1 = 1/f_3$. The fact that the ratio of buyers to sellers at the fair market (market 2) is unity follows by analogous reasoning.

Let us first calculate the value of the intensity of choice at which traders start to fragment weakly. To do so, for a given value of the free parameter θ_1, we start from low values of β and gradually increase the intensity of choice until it reaches a critical value where the single agent dynamics has two stable fixed points. Those values of β are shown by the upper solid line in figure 3.

The natural continuation of this analysis is to look—if it exists—for the strong fragmentation threshold. While thanks to our previous analysis of symmetric markets we know that for $\theta_1 = 0.5$ strong fragmentation takes place at $\beta = 1/0.252$, our numerical methods show that for reasonably asymmetric markets, i.e. $\theta_1 < 0.48$, strong fragmentation does not take place across the entire range of values of β that we consider numerically for our phase diagram. For θ_1 between 0.48 and 0.5, our numerics suggest possible strong fragmentation but a definite conclusion cannot be reached given the numerical precision limits of the required action minimizations.

To distinguish between different types of steady states in the following analysis—the number of emergent loyalty groups, their market preferences and sizes, we now introduce a triangle notation that is illustrated in figure 2 and used in the $(\theta_1, 1/\beta)$ phase diagram there. Each of the triangle corners represent preferences for one of the three markets, while full and empty circles represent large/small peaks; different colours denote the different trader classes. This notation allows us to

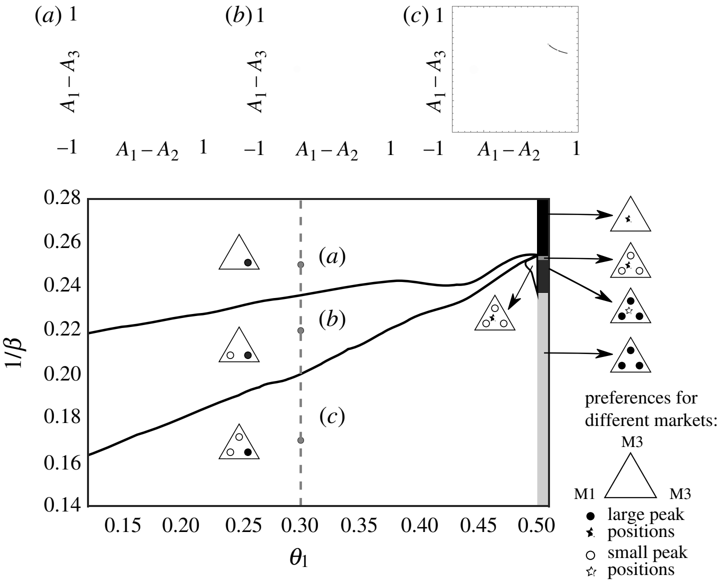

Figure 3. Peak structure of the steady-state distribution of traders' preferences when they learn to choose between three markets, two of which have symmetric market biases $\theta_1 = 1 - \theta_3$ and one of which is fair. The three insets on the top show the fixed point structure for an agent from class 2 ($p_{\mathcal{B}}^{(2)} = 0.2$), for $\theta_1 = 0.3$ and different β as indicated in the phase diagram by grey points. Full circles in the phase diagram correspond to a stable (large peak) and empty circles to a metastable fixed point (small peak), colours (black = class 1, red = class 2) differentiate between the agent classes. The grey band at $\theta_1 = 0.5$ shows the type of attraction distribution for the case $\theta_1 = 0.5$, i.e. when the three markets are fair; examples of attraction distribution peak structures in that region are shown in figure 2.

quickly realize whether some markets lost the competition, which markets are dominant, and which might attract only a single class of traders. Additionally, we use a star to denote an attraction distribution peak without preferences for a specific market. This is present only for the scenario with three fair markets, as depicted in the right band of the phase diagram in figure 4. The triangular representations shown on the right correspond to the flow diagrams with fixed points depicted in figure 2.

In figure 3, we see that for any value of β and $\theta_1 < 0.5$, the majority of the traders will prefer to trade at the fair market (market number two), so that this market will have a monopoly in the $r \rightarrow 0$ limit. When agents have finite memory, all three markets coexist when β is greater than the weak fragmentation threshold, but market 2 still attracts the majority of trades. Interestingly, in the region of the phase diagram with intermediate β (see inset (*b*)), all three markets coexist, but markets 1 and 3 are visited by only a single class, despite the fact that trading opportunities are lower that way.

In summary, the results depicted in figure 3 tell us that, apart from the particular case when the three markets are all fair, *strong* fragmentation does not take place when a fair market competes against two symmetrically biased markets. We therefore move next to an even less symmetric situation.

We continue exploration of the space of market biases by considering two symmetric markets with fixed market biases $\theta_1 = 0.3$ and $\theta_3 = 0.7$; this is the market set-up we mostly studied in previous works. Without the third market, when the two classes of traders adaptively choose between two symmetric markets one finds both weak and strong fragmentation above $\beta_c = 1/0.28$ [8]. Here, we add the third market and vary its bias, which as figure 4 shows leads to a range of different steady-state attraction distributions.

We first note that strong fragmentation appears, and does so across a reasonably broad range of market biases (grey zone in figure 4). This range excludes the case studied above where market 2 is fair: strong fragmentation occurs only for $\theta_2 \notin [0.45, 0.55]$, i.e. when the second market is sufficiently biased. For $\theta_2 < 0.45$ (resp. $\theta_2 > 0.55$) the traders from the first (resp. second) class *strongly fragment* across the two markets that maximize average profit per trade for each class. For example, in the case of $\theta_2 = 0.4$, buyers (traders in class 1, who have $p_{\mathcal{B}}^{(1)} = 0.8$) will prefer trading at markets 1 and 2 while the sellers remain unfragmented.

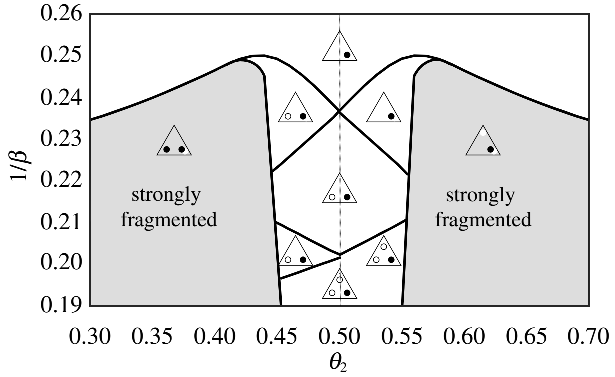

Figure 4. Types of attraction distributions in the population choosing between markets $\theta_1 = 1 - \theta_3 = 0.3$ and varying θ_2. The grey zone indicates the region in parameter space where the distribution of attractions has two large peaks for at least one class of agents, i.e. where strong fragmentation occurs. Note that between every unfragmented and strongly fragmented region (appearance of large loyalty groups at market 1 and 3) there is always a weakly fragmented region (where the same loyalty group, i.e. peak in the distribution, is small), but these regions are mostly too narrow to be visible. The grey line in the centre corresponds to the dashed line in figure 3.

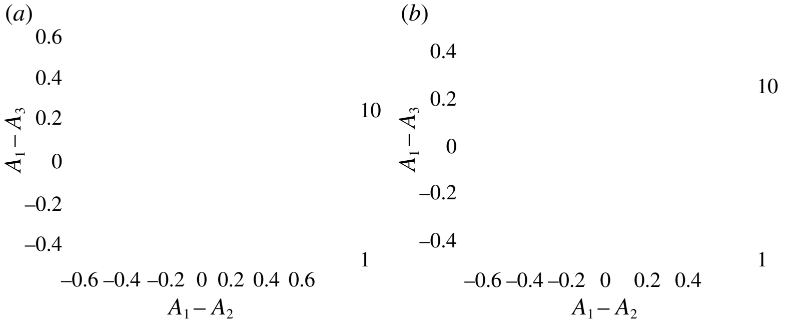

Figure 5. Distribution of attraction differences of traders who choose between three markets with market biases $(\theta_1, \theta_2, \theta_3) = (0.3, 0.35, 0.7)$. The population consists of two classes of $N/2 = 10^4$ traders with symmetric buy-sell preferences $p_{\mathcal{B}}^{(1)} = 1 - p_{\mathcal{B}}^{(2)} = 0.8$, inverse memory length $r = 0.01$ and intensity of choice $\beta = 1/0.21$. We see that the attraction distribution of the first class is strongly fragmented (panel (*a*)), while the second one is unfragmented (panel (*b*)), as predicted by the phase diagram in figure 4.

We do not explore the phase diagram below the first strong fragmentation threshold as this would require the numerical solution of self-consistency conditions for multiple aggregates in the presence of two (or more) strong fragmentation peaks in the traders' attraction distributions. This is numerically very challenging and so we leave it for future work. However, it is possible to get an intuition about the shape of the phase diagram below this threshold by extrapolating the zones of weak fragmentation in the range of θ_2 where the second market is close to fair.

We show in figure 4 graphically the types of steady state attraction distribution within the different regions of the phase diagram. These predictions are obtained using single agent flow diagrams as shown in figures 2 and 3. We show an exemplary comparison to stochastic multi-agent simulations in figure 5 and find excellent qualitative agreement. The agent class that mostly buys (class 1, left panel) fragments into two subpopulations mainly trading at markets 1 and 2, respectively, where they maximize their profit because θ_1, $\theta_2 < 0.5$. Agents in the class that mostly sells prefer market 2 as the less biased of the two markets that are populated by the buyers. We conjecture that it is the asymmetry imposed by two markets favouring buyers that leads to a consolidation around markets favouring buyers, while sellers do not develop attractions toward the market that favours them.

Having described the range of values of θ_2 for which strong fragmentation takes place, we inspect more closely the range of parameters for which only weak fragmentation occurs (figure 4). To do so, we look at how the attraction distributions of both classes of traders evolve at fixed $\theta_2 = 0.47$ when β increases. For values of β small enough in relation to the agents' attractions, they will essentially randomize their market choice, with a weak preference towards the market that is closest to fair, market 2. This preference increases with β so that traders from the two classes effectively coordinate at market 2, providing a good trade-off between profit and trading volume. As β grows further, additional small peaks arise in the attraction distributions while most of the traders remain in the fairer market. In particular, at $\beta = 1/0.246$ a peak corresponding to the strategy 'trading at the profit maximizing market' (market 1, which has $\theta_1 = 0.3$) appears for class 1. Then at $\beta = 1/0.228$, a peak corresponding to the strategy 'trading at the profit maximizing market' (market 3 with $\theta_3 = 0.7$) appears in the attraction distribution of the agents from the second class. After those two successive appearances of weak fragmentation between the fairer market and the profit maximizing market for both class 1 and class 2, further peaks in the attraction distribution—which correspond to the strategy 'trading at the volume maximizing market'—appear successively for class 2 at $\beta = 1/0.207$ and then for class 1 at $\beta = 1/0.198$.

Our phase diagram suggests that fairness of the second market weakens fragmentation. We cannot exclude, however, that strong fragmentation might occur even for θ_2 close to 0.5, for larger β (lower $1/\beta$) than investigated in the phase diagram of figure 4.

Interestingly, addition of the third market leads to trade shifting away from one of the symmetric markets, throughout the entire strong fragmentation region in figure 4. Only when the added market is close to fair can the two symmetric markets continue to coexist, though with both receiving only a small fraction of trades. Market 2 in fact has the largest market share throughout figure 4.

We can summarize the intuition behind the above results as follows. As the intensity of choice increases, each class of agents will first fragment weakly between a market that is close to fair (market 2) and the market that maximizes profit for them, and then fragment weakly across all three markets. On the other hand, if the second market is not fair, the class for which this market is more profitable will fragment strongly between their two profit maximizing markets, while the other class will only trade at the market that is closest to fair. The results of this subsection suggest that as soon as traders have at their disposal a reasonably fair market, they are not going to fragment and will prefer to trade with the fair market; when they have no fair market they will always prefer the profit maximizing market, and will visit the volume maximizing market (which brings lower profits but typically more trades) only as a last resort.

The two examples presented in §§4.1 and 4.2 lead to the conjecture that the presence of a fair or nearly fair market—which provides a good trade-off between profit in individual trades and trading volume—can suppress fragmentation. To confirm this conjecture, we consider three markets where the first one is biased toward buyers ($\theta_1 = 0.3$) and the second one is fair ($\theta_2 = 0.5$); the bias of the third market is the parameter we will vary.

As we did in the previous subsections, we will draw a phase diagram of the type of attraction distribution for the two agent classes, as a function of the intensity of choice β and the bias of the third market $\theta_3 \in [0, 1]$. The result in figure 6 shows that within the range of parameters explored, if there is fragmentation it is weak, so that the attraction distributions for both trader classes always become unimodal in the $r \to 0$ limit. (Extrapolation to lower $1/\beta$ than shown in figure 6 suggests that this situation does not change at even larger intensity of choice.) Only one peak has weight of order one and, depending on the values of β and θ_3, the steady state is either unfragmented or weakly fragmented, having one or two small peaks that disappear in the $r \to 0$ limit.

One notes that once the intensity of choice increases above a certain threshold value shown by the full black line in figure 6, a weak peak corresponding to the strategy 'trading at market 1' appears in the distribution of attractions of the first class of agents, whose attractions are marked by black circles; recall here that market 1 provides buyers, who are more frequent among agents of the first class, with higher returns. When β crosses the second fragmentation threshold (red line in figure 6), the same type of weak peak emerges in the distribution of attractions of the second class of agents (denoted by a red empty circle as before).

The fact that the two solid lines just described are close to horizontal reflects the fact that since almost all of the population trades at the fair market, the bias of the third market will not significantly influence the preference of traders. This is the reason why the intensity of choice at which traders of class 1 (resp. class 2) will weakly fragment between markets 1 and 2 is almost independent of the bias of the third market. The same is not true of the thresholds for the appearance of a peak corresponding to the strategy 'trade at market 3', which are indicated by the sloping dashed lines in figure 6.

Consistent with previously discussed results, the existence of fair market suppresses strong fragmentation and within the space of parameters depicted in figure 6 we note only weak

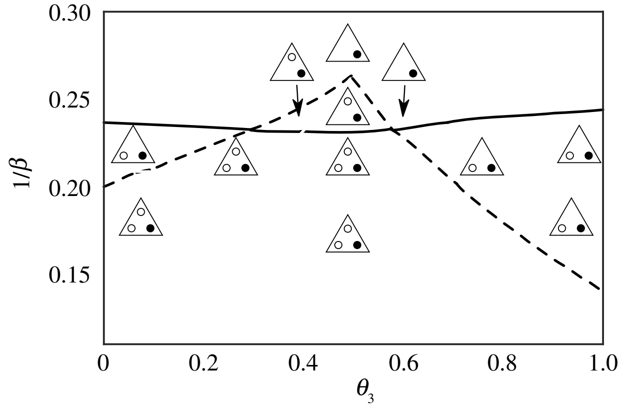

Figure 6. Peak structure of the different attraction distributions when $\theta_1 = 0.3$, $\theta_2 = 0.5$, $p_{\mathcal{B}}^{(1)} = 1 - p_{\mathcal{B}}^{(2)} = 0.8$. The solid/dashed lines show weak fragmentation transitions where subpopulations emerge that favour markets 1 or 3 (line colours denote class of agents in which transition occurs).

fragmentation. This means that across the parameter range investigated the fair market attracts most of the traders. We also note that the third market loses the competition when it is very biased and the intensity of choice is not large enough (note regions where market 3 either attracts none or only one class). However, it is interesting to see that for sufficiently large intensity of choice β all three markets coexist independently of the third market bias.

So far we have discussed various cases of fragmentation in the three-market set-up. We found that above some critical value of the intensity of choice β, the solution in which the population remains indecisive towards the markets is never stable and at least one market loyalty group is formed. The obvious question is now whether we can say something about the number of distinct agent subpopulations in the general case of M markets.

The theoretical description of the population's adaptation in the most general case, without market or agent symmetry requires the self-consistent procedure of calculating order parameters (one per market) and the steady-state distribution of the agent attractions. This is a non-trivial task in higher dimensions but the general existence of solutions can be rationalized within a simple counting argument.

In the following, we make the assumption that for all M there is a fragmentation threshold β_s above which the drift in the Fokker–Planck representation of the dynamics has multiple zeros. However, even when this is the case it is not clear whether all agent classes will develop loyalty groups towards each of the markets (and the corresponding attraction distribution peaks), whether the peaks will be small or large; in the latter case fragmentation persists by definition in the $r \to 0$ limit. To address this question we consider an agent class that is strongly fragmented across M markets so that in the limit $r \to 0$ its attraction distribution consists of M delta peaks with weights of order unity. We can find the peak positions by locating the zeros of the drift, but without the Fokker–Planck solution, we cannot obtain the peak weights and the Freidlin–Wentzell approach becomes difficult. We therefore ask how many non-zero peak weights can exist in general, for C agent classes and M markets. As explained, we assume the general shape of the steady-state distribution

$$P^{(c)}(\mathbf{A}) = \sum_{m=1}^{M} \omega_m^{(c)} \delta(\mathbf{A} - \mathbf{A}_m^{(c)}).$$

Each of the agent classes is described by peak weights $\omega_1^{(c)}, \ldots, \omega_M^{(c)}$ that satisfy the normalization condition $\sum_{m=1}^{M} \omega_m^{(c)} = 1$, thus in the absence of any symmetry we have $M-1$ free variables per class. On the other hand, for each market we define an order parameter f_m, thus the system of equations we need to solve to find a strongly fragmented solution is

$$\mathcal{F}_m(\omega_1^{(1)}, \omega_2^{(1)}, \ldots, \omega_M^{(1)}, \ldots, \omega_M^{(C)}) = f_m.$$

Here $\mathcal{F}_m$ denotes the relationship between the peak weights and market order parameters; an example of this for $C = 2$ and $M = 3$ is written explicitly in equation (4.1). Without symmetries, when all the equations and variables are independent, this system of M equations and $C(M - 1)$ variables has a unique solution only when the number of equations is equal to the number of variables, i.e. $M = C(M - 1)$. This equation has an integer solution pair only when both number of market M and classes C is equal to two, $(C, M) = (2, 2)$. For example, the population studied so far with its $C = 2$ agent classes requires $2(M - 1)$ weights for strong fragmentation across M markets, and equating the number of variables $2(M - 1)$ and the number of equations M gives $M = 2$ markets, which is the case studied in [6].

Since we have seen that full fragmentation, with all agent classes developing separate loyalty groups for all markets, can only happen (without symmetries) in systems with two markets and two agent classes, we next relax the assumption on the number of loyalty groups. Let us suppose there are M markets and two agent classes, each of them fragmenting into $\eta^{(c)}$ subgroups (i.e. having only $\eta^{(c)}$ non-zero peak weights), the system of equations for these weights has a unique solution when $\eta^{(1)} + \eta^{(2)} - 2 = M$. This shows that if one class divides into M loyalty groups, the second class will fragment only across two markets; other combinations satisfying $\eta^{(1)} + \eta^{(2)} = M + 2$ are also possible. For a general number of agent classes, the analogous constraint reads

$$\eta^{(1)} + \eta^{(2)} + \cdots + \eta^{(C)} = M + C. \tag{5.1}$$

As an example, if one class develops loyalty groups to all M markets, the other $C - 1$ classes can have C such subpopulations in total, equating to one bimodal and $C - 2$ unimodal steady-state distributions. More generally, if we associate each loyalty group with its preferred market then (5.1) shows that it is impossible for the population classes to develop disjoint sets of preferred markets, as that would require $\eta^{(1)} + \eta^{(2)} + \cdots + \eta^{(C)} \leq M$. For example, in the case $C = 2$, there will be at least two markets for which both classes have loyalty groups; the overlap will be even greater if some markets lose out and have no associated loyalty group.

Summarizing, the conclusion of our counting argument is that in the $r \to 0$ limit at most $C + M$ loyalty groups can coexist. In the three-market scenario with two classes, this is at most five loyalty groups. We saw an exception in the case of three fair markets, where six loyalty groups can exist; this is because of the symmetry between the markets, which our general argument excludes. It is remarkable how the simple counting argument gives a variety of new conjectures for the systems with multiple markets. It provides a maximal number of loyalty groups; it tells us that all markets can in principle coexist, and that the loyalty groups of different agent classes must overlap at C markets at least. An interesting consequence is the emergence of a state where some markets are persistently visited only by a subset of the overall population of traders.

In this paper, we have investigated whether market coexistence is possible in systems with more than two markets when agents with fixed buy/sell preferences adapt dynamically to optimize their choice of market. This research question is motivated by empirical observations of multiple markets coexisting and attracting loyal traders both in *in silico* and real market competitions. Rather than aiming to reproduce market stylized facts, here we investigate mechanisms that might lead to a previously neglected phenomenon, namely, that multiple market loyalties, and thus market coexistence, could emerge without any underlying heterogeneity of agents or markets and only as a consequence of the co-adaptation of the agents. To this end, we studied the possible steady states of the agent dynamics, in particular with regard to the occurrence of fragmentation, where a homogeneous class of agents spontaneously forms subpopulations with long-lived market preferences.

The proposed model contains an implicit assumption of bounded rationality as the agents do not optimize any utility function or aim to make the rational/optimal choice; instead their behaviour is based on their past observed outcomes. Depending on the learning parameters the agents are tunable between trading randomly and a behaviour that repeats the most rewarding past choices. The agents do not possess knowledge about market mechanisms nor the existence of various different agents nor their scores, they only make decisions based on their past observations. In this regard, these assumptions violate rational agent assumptions due to the lack of information and lack of utility-optimizing behaviour. Nonetheless, in the case of two markets it has been shown [7] that when the agents' memory is infinitely long ($r \to 0$) and they do not update their preferences for options they did not try in the last steps, then the expected outcome under rational behaviour (Nash equilibrium) is retrieved.

Motivated by the wide variety of structures of the attraction distributions that one observes in multi-agent simulations, we explored different combinations of market biases and their influence on the phenomenon of fragmentation. First we studied fragmentation across three fair markets, i.e. with $\theta_1 = \theta_2 = \theta_3 = 0.5$. This was the only scenario where we found that all three markets coexist across the full range of the intensity of choice β of the agents. As β increases we nonetheless see a change, from an indecisive population (where agents visit all three markets randomly) to a strongly fragmented population where each agent class splits into three equal-sized loyalty groups with a distinct preference for one market.

We continued by exploring different market configurations to get an intuition for the factors that drive fragmentation. This enabled us to identify two principal causes of fragmentation: (i) the *similarity between the markets' biases*, (ii) the *average volume of trade and average profit earned at a market*. The *similarity* between two markets is going to enhance fragmentation because traders are more likely to split across two markets if they effectively cannot tell them apart. This effect is visible in §4.3 where the strong and weak fragmentation thresholds are the highest (in terms of $1/\beta$) when the second market and the fair market have the same bias. The ordering of the appearance of the peaks in the traders' attraction distributions suggests—as we pointed out in §4.2—that traders will have an initial preference for markets that provide a good balance between trading volume and profit, then as the intensity of choice increases they will first spread to the market that maximizes their profit and then subsequently to the one that maximizes their trading volume.

The concepts of positive and negative size effects introduced previously [17,18] are useful when thinking about traders who develop loyalty for markets that do not reward them highly. At these markets, traders benefit from the many trading options available (positive size effects), and the fact that they are in the minority group (negative size effects). However, contrary to the findings of Ellison *et al.* [17] and Shi *et al.* [18], we note that market coexistence is more prevalent when the markets are similar—the fragmentation region shrinks with increased market difference.

Apart from the case of three identical markets, we find that once β is large enough for agents to stop choosing markets at random, the three markets never coexist fully in the large memory limit, i.e. at least one of them will have a market share that vanishes for $r \to 0$. At most, we observe that the population fragments strongly across *two* markets (see strong fragmentation in figure 4). These markets then each have a finite share of the trading volume for $r \to 0$, though with one being subdominant because it is visited only by (some of the) agents from a single class.

From a general counting argument, we found further that full market coexistence, where all agent classes develop the (joint) maximal number of loyalty groups, leads to apparently specialized markets: some agent classes develop loyalties only to a subset of all markets (as in figure 4) and conversely some markets are not visited by agents from all classes. This is not a consequence of a market explicitly targeting some subset of the agent population, but rather of the limited number of market loyalties the different agent classes can support.

We mostly considered moderate values of β driven by our interest in finding domains of different steady states, and for those purposes our straightforward implementation of the action minimization algorithm served us well. However, for large values of β it occasionally fails to find minimal action path, thus robustness and accuracy improvements are needed if one is interested particularly in this regime. One possibility might be to use the geometric minimum action method [26].

Although the analytical and numerical methodology we have proposed to study agents who choose between multiple markets is valid for any number of markets M, it is challenging for two reasons: (i) the parameter space dimension grows with M thus making numerical exploration of all possible behaviours difficult, and (ii) analytical approaches also become harder to implement as the analysis is done in the space of attraction differences of dimension $M - 1$.

Turning to implications for market competition, our results show that loyalty groups for all three markets rarely exist for large intensity of choice β in the large memory limit ($r \to 0$). However, for finite memory ($r > 0$), one should expect that the small peaks persist. In two market systems, above certain values of r (effectively for short memory) only a strongly fragmented steady state exists [6] instead of two weakly fragmented and metastable strongly fragmented states; it would be interesting to investigate if similar results also hold for multiple markets.

In this and previous studies, we have investigated how agents adapt based on their exploration of markets; the adaptation mechanism implicitly assumes that markets do not change. Realistically, one would expect that a market tries to adapt as well once the number of traders using it decreases. If markets only try to maximize this number of traders, one could speculate that by adapting their θ biases they would converge to all-fair markets (similarly to the Hotelling paradox [27]). If on the other hand markets were to adapt to optimize the number of *successful* trades, by e.g. charging fixed or profit-dependent fees, then it would be intriguing to know what types of steady states would be realized in the overall system of agents and markets.

Finally, a broad implication of our study is that fragmentation (weak or strong) can emerge spontaneously within a class of homogeneous traders, in contrast to statements elsewhere [1] arguing that heterogeneity among traders is the reason for market fragmentation. This we think is a very interesting result as it demonstrates that structure in the preferences of economic agents might emerge out of adaptation rather than being present from the start. To this end, we made an assumption of homogeneity of agents in terms of their learning parameters, which simplified the mathematical description but could be relaxed and investigated further. Heterogeneity in agents' memory parameter r was investigated in [28] where it was shown that a population containing both fast ($r = 1$) and slow ($r \ll 1$) agents still fragments across two markets, with the critical β depending on the fraction of fast traders. Heterogeneity in β might be mathematically more challenging but could in principle be tackled following the procedures outlined in [8]. The population can be split into subgroups of traders with the same β whose steady-state market preference distributions should be found assuming fixed demand-to-supply market parameters. Finally, it should be checked whether those market aggregated parameters can be reproduced from the trader preferences obtained, i.e. whether the overall solution is self-consistent. This would be an interesting next step to investigate, together with heterogeneities in terms of trading strategies and budget constraints.

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.cz8w9gj2n [29].

We declare we have no competing interests.

R.N., A.A. and P.S. conceptualized the model and formalized the mathematical framework. R.N. implemented the numerical simulations and prepared the figures. R.N., A.A. and P.S. analysed the results and discussed their implications. All authors wrote the manuscript and gave final approval for publication.

No funding has been received for this article.

The authors are grateful to Peter McBurney for useful discussions. P.S. acknowledges the stimulating research environment provided by the EPSRC Centre for Doctoral Training in Cross-Disciplinary Approaches to Non-Equilibrium Systems (CANES, EP/L015854/1). A.A. acknowledges the funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science, and Technological Development of the Republic of Serbia.

In this appendix, we give the expression of the drift and covariance matrix that appear in the Kramers–Moyal expansion in equation (3.1). We only give the results here; the steps of the derivation can be found in the thesis of Alorić [29]. First, the drifts of the attraction differences are

$$\mu_2^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3) = \Big(\mathcal{P}_1^{(c)}(f_1)P(M=1) - \mathcal{P}_2^{(c)}(f_2)P(M=2)\Big) - \Delta A_2^{(c)} \tag{A 1}$$

and

$$\mu_3^{(c)}(\mathbf{\Delta A}^{(c)}, f_1, f_2, f_3) = \Big(\mathcal{P}_1^{(c)}(f_1)P(M=1) - \mathcal{P}_3^{(c)}(f_3)P(M=3)\Big) - \Delta A_3^{(c)}. \tag{A 2}$$

Here $\mathcal{P}_m^{(c)}(f_m)$ is the average payoff of a trader from class c at market m and $P(M = m)$ is the probability to trade at market m, which depends on the vector $\Delta A^{(c)}$ of attraction differences. We do not write this dependence explicitly to lighten the notations. The f_m are the market aggregates, i.e. buyer-to-seller ratios, at the three markets. In order to check the validity of our calculations we compared the dynamics of the aggregate f_1 during a multi-agent simulation with the evolution of the aggregates under the homogeneous population dynamics as detailed in [7], finding good agreement as shown in figure 7.

We next look at the covariance matrix of the effective noise acting on the attraction differences

$$\begin{pmatrix} \Sigma_{22}^{(c)} & \Sigma_{23}^{(c)} \\ \Sigma_{23}^{(c)} & \Sigma_{33}^{(c)} \end{pmatrix}, \tag{A 3}$$

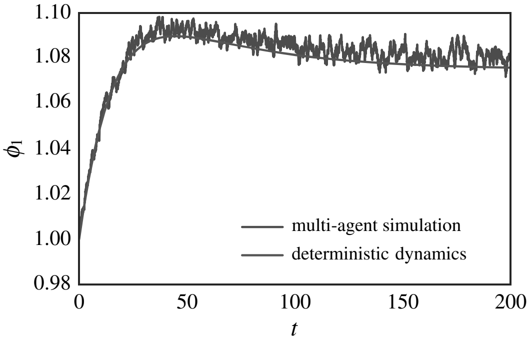

Figure 7. Comparison between the time series of the aggregate (ratio of buyers to sellers) at the first market during a multi-agent simulation (with $r = 0.01$ and 10^4 agents in each class) and its evolution under the homogeneous population dynamics. The parameters for the plots in this figure are $(\theta_1, \theta_2, \theta_3) = (0.2, 0.5, 0.8)$, $\beta = 1/0.3$ and $p^{(1)}_{\mathcal{B}} = 1 - p^{(2)}_{\mathcal{B}} = 0.8$.

which is given by

$$\begin{aligned}\Sigma^{(c)}_{22}(\Delta\mathbf{A}^{(c)}, f_1, f_2, f_3) = {} & \left(\mathcal{Q}^{(c)}_1(f_1) - 2\Delta A^{(c)}_2 \mathcal{P}^{(c)}_1(f_1)\right)P(M=1) \\ & + \left(\mathcal{Q}^{(c)}_2(f_2) - 2\Delta A^{(c)}_2 \mathcal{P}^{(c)}_2(f_2)\right)P(M=2) + \Delta A^{(c)^2}_2,\end{aligned} \tag{A 4}$$

$$\begin{aligned}\Sigma^{(c)}_{33}(\Delta\mathbf{A}^{(c)}, f_1, f_2, f_3) = {} & \left(\mathcal{Q}^{(c)}_1(f_1) - 2\Delta A^{(c)}_3 \mathcal{P}^{(c)}_1(f_1)\right)P(M=1) \\ & + \left(\mathcal{Q}^{(c)}_3(f_3) - 2\Delta A^{(c)}_3 \mathcal{P}^{(c)}_3(f_3)\right)P(M=3) \\ & + \Delta A^{(c)^2}_3\end{aligned} \tag{A 5}$$

and

$$\begin{aligned}\Sigma^{(c)}_{23}(\Delta\mathbf{A}^{(c)}, f_1, f_2, f_3) = {} & \Delta A^{(c)}_2 \left(P(M=3)\mathcal{P}^{(c)}_3(f_3) - P(M=1)\mathcal{P}^{(c)}_1(f_1)\right) \\ & + \Delta A^{(c)}_3 \left(P(M=2)\mathcal{P}^{(c)}_2(f_2) - P(M=1)\mathcal{P}^{(c)}_1(f_1)\right) \\ & + P(M=1)\mathcal{Q}^{(c)}_1(f_1) + \Delta A^{(c)}_2 \Delta A^{(c)}_3,\end{aligned} \tag{A 6}$$

where $\mathcal{Q}^{(c)}_m(f_m)$ is the average squared payoff, see [7].

We describe in this section the large deviation methods we use to study multimodal attraction distributions in the steady state of our agents' learning dynamics. As explained in more detail in [7], steady-state attraction distributions for small r will be peaked around the stable fixed points of the single agent dynamics. The shape of these peaks becomes Gaussian for $r \to 0$, with a covariance matrix proportional to r that is straightforward to determine. Much more difficult to find are the *weights* of the peaks as these involve rare fluctuations of an agent making the transition from one peak to another. In one dimension, the problem is tractable as an explicit formula for the steady-state distribution of attractions can be given [6]. In higher dimensions detailed balance [31] would have a similar simplifying effect, but our single agent dynamics in the two-dimensional attraction space (for each class of agents) does not have this property.

In our approach, we therefore consider the peak weights in an attraction distribution as a result of the balance between transitions between the various peaks. We therefore need to find the rates for these transitions. To do this, note from the Kramers–Moyal expansion that the single agent learning is described by a Langevin equation with noise variance $O(r)$. For $r \to 0$, we are therefore looking for transition rates in a low noise limit. This allows us to use Freidlin–Wentzell theory, which deals with large deviations of Langevin dynamics in exactly this limit [32].

B.1. Freidlin–Wentzell theory

We use Freidlin–Wentzell theory in the form developed in [33,34], which generalizes the Eyring–Kramers [35] formula for the rates of noise-activated transitions to non-conservative dynamics. We give a brief summary of those aspects of Freidlin–Wentzell theory that we use in our numerical application and refer to [32] for a mathematically rigorous description and to [33] for a more statistical physics-oriented summary.

Freidlin–Wentzell theory is concerned with the transition rates between two stable states (here $A_1^\star$ and $A_2^\star$; below we drop the Δ from the notation for the attraction differences for brevity) of a non-conservative stochastic dynamics in the low noise limit. A general Langevin equation can be written in the form

$$\dot{A}(t) = \boldsymbol{\mu}(\mathbf{A}(t)) + \sqrt{r}[\Sigma(\mathbf{A}(t))]^{1/2}\boldsymbol{\xi}(t), \tag{B 1}$$

where $\boldsymbol{\xi}(t)$ is white noise with unit covariance matrix. The drift $\boldsymbol{\mu}$ and the covariance matrix Σ of the noise in the Langevin equation are given in [7] for our learning dynamics. In the generic version above, we have omitted the superscript (c) indicating the class of agents we are considering, as well as the dependence of drift and noise covariance on the market aggregates.

Associated with the Langevin dynamics is an Onsager–Machlup action $\mathcal{S}[A]$ for any path $A(t)$

$$\mathcal{S}[A] = \int_{t_1}^{t_2} \frac{1}{2}\Big(\dot{\mathbf{A}}(t) - \boldsymbol{\mu}(\mathbf{A}(t))\Big)^T \Sigma^{-1}(\mathbf{A}(t))\Big(\dot{\mathbf{A}}(t) - \boldsymbol{\mu}(\mathbf{A}(t))\Big)\,\mathrm{d}t. \tag{B 2}$$

The action determines the probability of observing any path $[A(t)]$ according to

$$\Gamma_{1\to 2} \sim \exp\left(-\frac{\mathcal{S}[\mathbf{A}]}{r}\right), \tag{B 3}$$

where $\sim$ means that the equality is true up to a prefactor (which depends on the time discretization used). The main Freidlin–Wentzell result we need is that the rate $\Gamma_{1\to 2}$ for a transition from $A_1^\star$ to $A_2^\star$ (*forward path*) is [32,36]

$$\Gamma_{1\to 2} \sim \exp\left(-\frac{\mathcal{S}^\star_{1\to 2}}{r}\right), \tag{B 4}$$

where $\mathcal{S}^\star_{1\to 2}$ is the minimal action achievable by any path from $A_1^\star$ to $A_2^\star$ in the infinite time interval $(t_1, t_2) = (-\infty, \infty)$. The rate $\Gamma_{2\to 1}$ for the *reverse* transition from $A_2^\star$ to $A_1^\star$ is similarly $\Gamma_{2\to 1} \sim \exp(-\mathcal{S}^\star_{2\to 1}/r)$.

The attraction distributions we are after will consist of narrow (for small r) peaks at $A_1^\star$ and $A_2^\star$. The weights ω_1 and ω_2 of these two peaks, which represent the probability for an agent to be within each peak, must then be such that forward and backward transitions balance

$$\omega_1\Gamma_{1\to 2} = \omega_2\Gamma_{2\to 1} \tag{B 5}$$

and

$$\frac{\omega_1}{\omega_2} \propto \exp\left(\frac{\mathcal{S}^\star_{1\to 2} - \mathcal{S}^\star_{2\to 1}}{r}\right). \tag{B 6}$$

This expression shows that when the forward and backward minimal actions are not equal, then one of the two peaks will have an exponentially small weight as $r \to 0$. In practice, this is true when the action difference inside the exponential in (B 5) is large compared with r. If it is only of order r or smaller, then we cannot say anything about the weights as we do not determine the prefactor in (B 5), though we would expect them to be of order unity.

B.2. Finding the minimal action path numerically

Following the method of Bunin *et al.* [36], we find the minimal action by discretizing the path $[A(t)]$, evaluating the action as a function of this discretized path and then minimizing with respect to the (discretized) path. The path is discretized into 10 equally spaced time steps between $t = 0$ and $t = 10$; we found this choice of parameters to be a reasonable trade-off between the precision of our result and the complexity of minimizing the discretized action.

There are other methods for finding the minimal value of the action defined in equation (B 2), such as solving a Hamilton–Jacobi equation [33], but we chose to use the path discretization method because we found this to be more robust with respect to changes of model parameters. The discretization approach could also be improved further, using for example the geometric minimum action method [26], but we found that this was not necessary to achieve the desired precision. We tested this e.g. by benchmarking against closed-form results that can be obtained for $M = 2$ [6].

The numerical path optimization can be simplified by restricting attention to the *activation* part of the path. Generally, for a system with two stable fixed points $A_1^\star$ and $A_2^\star$ and one saddle point $\bar{A}$ between them, the optimal path starting from $A_1^\star$ will pass through the saddle point $\bar{A}$ and then relax to $A_2^\star$ following the relaxation dynamics $\dot{A}(t) = \boldsymbol{\mu}(A(t))$, equation (B 2) shows that the relaxation dynamics does not contribute to the total action as the integrand (the Lagrangian) vanishes identically along this section of the path. As a consequence, the problem of finding a minimal action path between $A_1^\star$ and $A_2^\star$ can be reduced to finding the minimal action path between $A_1^\star$ and $\bar{A}$, i.e. from the initial fixed point to the saddle. This restriction significantly improves the precision of the numerical path optimization.

With the above method, we can work out the action difference between any two fixed points of the single agent dynamics, as a function of the market aggregates. The values of these aggregates where the action difference between two single agent fixed points vanishes identify the points where the steady state attraction distribution of our learning can have more than one peak. Either side of these values, a single peak is dominant in the attraction distribution; which peak this is changes discontinuously at a zero action difference value of the market aggregates.

1. Gomber P, Sagade S, Theissen E, Weber MC, Westheide C. 2017 Competition between equity markets: a review of the consolidation versus fragmentation debate. *J. Econ. Surv.* **31**, 792–814. (doi:10.1111/joes.12176)
2. O'Hara M, Ye M. 2011 Is market fragmentation harming market quality? *J. Financ. Econ.* **100**, 459–474. (doi:10.1016/j.jfineco.2011.02.006)
3. Hasbrouck J. 1995 One security, many markets: determining the contributions to price discovery. *J. Finance* **50**, 1175–1199. (doi:10.1111/j.1540-6261.1995.tb04054.x)
4. Shorter G, Miller RS. 2014 Dark pools in equity trading: policy concerns and recent developments. Technical report. See https://digital.library.unt.edu/ark:/67531/metadc461960/.
5. Alorić A, Sollich P, McBurney P. 2015 Spontaneous segregation of agents across double auction markets. In *Advances in artificial economics* (eds Frédéric Amblard, Francisco J. Miguel, Adrien Blanchet, Benoit Gaudou), vol. 676. Lecture Notes in Economics and Mathematical Systems, pp. 79–90. Berlin, Germany: Springer International Publishing.
6. Alorić A, Sollich P, McBurney P, Galla T. 2016 Emergence of cooperative long-term market loyalty in double auction markets. *PLoS ONE* **11**, 1–26. (doi:10.1371/journal.pone.0154606)
7. Nicole R, Sollich P. 2018 Dynamical selection of Nash equilibria using reinforcement learning: emergence of heterogeneous mixed equilibria. *PLoS ONE* **13**, e0196577. (doi:10.1371/journal.pone.0196577)
8. Alorić A, Sollich P. 2019 Market fragmentation and market consolidation: multiple steady states in systems of adaptive traders choosing where to trade. *Phys. Rev. E* **99**, 062309. (doi:10.1103/PhysRevE.99.062309)
9. Cai K, Gerding E, McBurney P, Niu J, Parsons S, Phelps S. 2009 Overview of CAT: a market design competition. Technical report, Department of Computer Science, University of Liverpool. (http://www.csc.liv.ac.uk/research/techreports/tr2009/ulcs-09-005.pdf)
10. Niu J, Cai K, Parsons S, Gerding E, McBurney P, Moyaux T, Phelps S, Shield D. 2008 JCAT: a platform for the TAC market design competition. In *Proc. of the 7th Int. Joint Conf. on Autonomous Agents and Multiagent Systems*, pp. 1649–1650. See http://portal.acm.org/citation.cfm?id=1402747.
11. Cai K, Niu J, Parsons S. 2014 On the effects of competition between agent-based double auction markets. *Electron. Commer. Res. Appl.* **13**, 229–242. (doi:10.1016/j.elerap.2014.04.002)
12. Niu J, Cai K, Parsons S, Sklar E. 2007 Some preliminary results on competition between markets for automated traders. *AAAI-07 Workshop on Trading Agent*, pp. 19–26. See http://www.aaai.org/Papers/Workshops/2007/WS-07-13/WS07-13-003.pdf.
13. Miller T, Niu J. 2012 An assessment of strategies for choosing between competitive marketplaces. *Electron. Commer. Res. Appl.* **11**, 14–23. (doi:10.1016/j.elerap.2011.07.009)
14. Gode DK, Sunder S. 1993 Allocative efficiency of markets with zero-intelligence traders: market as a partial substitute for individual rationality. *J. Polit. Econ.* **101**, 119–137. (doi:10.1086/261868)
15. Cliff D, Bruten J. 1997 Zero is not enough: on the lower limit of agent intelligence for continuous double auction markets. Technical Report HPL-97-141, Hewlett-Packard Laboratories, Bristol, UK.
16. Tóth B, Scalas E, Huber J, Kirchler M. 2007 The value of information in a multi-agent market model – the luck of the uninformed. *Eur. Phys. J. B* **55**, 115–120. (doi:10.1140/epjb/e2007-00046-2)
17. Ellison G, Fudenberg D, Möbius M. 2004 Competing auctions. *J. Eur. Econ. Assoc.* **2**, 30–66. (doi:10.1162/154247604323015472)
18. Shi B, Gerding EH, Vytelingum P, Jennings NR. 2013 An equilibrium analysis of market selection strategies and fee strategies in competing double auction marketplaces. *Auton. Agents and Multi-Agent Syst.* **26**, 245–287. (doi:10.1007/s10458-011-9190-5)
19. Caillaud B, Jullien B. 2003 Chicken & egg: competition among intermediation service providers. *RAND J. Econ.* **34**, 309–328. (doi:10.2307/1593720)
20. Duffy J. 2006 Agent-based models and human subject experiments. *Handb. Comput. Econ.* **2**, 949–1011. (doi:10.1016/S1574-0021(05)02019-8)
21. Ladley D. 2012 Zero intelligence in economics and finance. *Knowl. Eng. Rev.* **27**, 273–286. (doi:10.1017/S0269888912000173)
22. Anufriev M, Arifovic J, Ledyard J, Panchenko V. 2013 Efficiency of continuous double auctions under individual evolutionary learning with full or limited information. *J. Evol. Econ.* **23**, 539–573. (doi:10.1007/s00191-011-0230-8)
23. Watkins CJCH, Dayan P. 1992 Q-learning. *Mach. Learn.* **8**, 279–292. (doi:10.1023/A:1022676722315)
24. Camerer C, Ho TH. 1999 Experience-weighted attraction learning in normal form games. *Econometrica* **67**, 827–874. (doi:10.1111/1468-0262.00054)
25. Ho TH, Camerer C, Chong J-K. 2007 Self-tuning experience weighted attraction learning in games. *J. Econ. Theory* **133**, 177–198. (doi:10.1016/j.jet.2005.12.008)
26. Heymann M, Vanden-Eijnden E. 2008 Pathways of maximum likelihood for rare events in nonequilibrium systems: application to nucleation in the presence of shear. *Phys. Rev. Lett.* **100**, 140601. (doi:10.1103/PhysRevLett.100.140601)
27. Hotelling H. 1929 Stability in competition. *Econ. J.* **39**, 41–57. (doi:10.2307/2224214)
28. Nicole R. 2017 Fluctuations and large deviations in game theoretical models. PhD thesis, King's College London, London, UK. (https://kclpure.kcl.ac.uk/portal/files/94142430/2017_Nicole_Robin_1345260_ethesis.pdf)
29. Alorić A, Nicole R, Sollich P. 2021 Data from: Fragmentation in trader preferences among multiple markets: market coexistence versus single market dominance. Dryad Digital Repository. (doi:10.5061/dryad.cz8w9gj2n)

30. Alorić A. 2017 Spontaneous segregation of adaptive agents in auctions. PhD thesis, King's College London, London, UK.
31. Risken H. 1984 *The Fokker–Planck equation*. Berlin, Germany: Springer.
32. Freidlin M, Wentzell A. 1998 *Random perturbations of dynamical systems*. Berlin, Germany: Springer.
33. Bouchet F, Reygner J. 2016 Generalisation of the Eyring–Kramers transition rate formula to irreversible diffusion processes. *Annales de l'Institut Henri Poincaré* **17**, 3499–3532. (doi:10.1007/s00023-016-0507-4)
34. Bradde S, Biroli G. 2012 The generalized Arrhenius law in out of equilibrium systems. (http://arxiv.org/abs/1204.6027)
35. Kramers HA. 1940 Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284–304. (doi:10.1016/S0031-8914(40)90098-2)
36. Bunin G, Kafri Y, Podolsky D. 2012 Large deviations in boundary-driven systems: numerical evaluation and effective large-scale behavior. *Europhys. Lett.* **99**, 20002. (doi:10.1209/0295-5075/99/20002)

EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE

Open Access

Sustainability of Stack Exchange Q&A communities: the role of trust

Ana Vranić[1*], Aleksandar Tomašević[2], Aleksandra Alorić[1,3] and Marija Mitrović Dankulov[1]

*Correspondence: anav@ipb.ac.rs
[1] Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, Belgrade, Serbia
Full list of author information is available at the end of the article

Abstract

Knowledge-sharing communities are fundamental elements of a knowledge-based society. Understanding how different factors influence their sustainability is of crucial importance. We explore the role of the social network structure and social trust in their sustainability. We analyze the early evolution of social networks in four pairs of active and closed Stack Exchange communities on topics of physics, astronomy, economics, and literature and use a dynamical reputation model to quantify the evolution of social trust in them. In addition, we study the evolution of two active communities on mathematics topics and two closed communities about startups and compare them with our main results. Active communities have higher local cohesiveness and develop stable, better-connected, trustworthy cores. The early emergence of a stable and trustworthy core may be crucial for sustainable knowledge-sharing communities.

Keywords: Networks structure; Dynamic reputation; Knowledge exchange; Stack Exchange; Sustainability of Q&A communities

1 Introduction

The development of a knowledge-based society is one of the critical processes in the modern world [1, 2]. In a knowledge-based society, knowledge is generated, shared, and made available to all members. It is a vital resource. Sharing this resource between individuals and organizations is a necessary process, and knowledge-sharing communities are one of the fundamental elements of a knowledge society.

Often, these knowledge-sharing communities depend on the willingness of their members to engage in an exchange of information and knowledge. Participation in the community is voluntary, with no noticeable material gains for members. Recent research has shown that the process of knowledge and information exchange is strongly influenced by *trust* [3, 4]. The exchange of knowledge depends on trust between a member and the community. It is a collective phenomenon that depends on and is built through social interactions between community members. This is why we believe it is crucial to understand how trustworthy knowledge-sharing communities emerge and disappear, as well as to unveil the fundamental mechanisms that underlie their evolution and determine their sustainability.

Unlike small offline knowledge-sharing groups, online communities consist of a large number of members where repeatable mutual interactions between all members are not possible. Thus, the trustworthiness of individuals in these communities has to be assessed and signaled using other means. It was shown that the reputation of an individual within the community is a strong signal of her trustworthiness that can override the main sources of social bias [5]. The reputation helps users manage the complexity of the collaborative environment by signaling out trustworthy members.

In the past two decades, we have witnessed the emergence of an online knowledge-sharing community Stack Overflow, which has become one of the most popular sites in the world and the primary knowledge resource for coding. The success of Stack Overflow led to the emergence of similar communities on various topics and formed the Stack Exchange (SE) network.[1] The advancement of Information and communication technologies (ICTs) have enabled faster and easier creation and sharing of knowledge, but also the access to a large amount of data that allowed a detailed study of their emergence and evolution [6], as well as user roles [7], and patterns of their activity [8–10]. However, relatively little attention has been paid to the sustainability of SE communities. Most research focused on the activity and factors that influence the users' activity in these communities. Factors such as the need for experts and the quality of their contributions have been thoroughly investigated [11]. It was shown that the growth of communities and mechanisms that drive it might depend on the topic around which the community was created [12].

In this paper, we investigate the role of network structure and social trust dynamical user reputation in the sustainability of a knowledge-sharing community. Research on the sustainability of social groups shows that social interaction and their structure influence the dynamics and sustainability of social groups [13–16]. Due to large number of users and the smaller probability of repeated interactions dyadic trust between members may not play an essential role in the group dynamics of knowledge-sharing communities. However, it is known that the reputation of users, one of the proxies of trust in online communities, is the primary for them to become and maintain their productive member status [17–19].

With the proliferation of misinformed decisions, it is crucial to understand how to foster communities that promote collaborative knowledge exchange and understand how cooperative norms of trustworthy behavior emerge. The way people interact, specifically the structure of their interactions [20], and how inclusive and trustworthy the key members of the community can influence the sustainability of the knowledge-sharing communities. Although the topic and early adopters are essential in establishing a new SE community, they are not sufficient for sustainability. The current SE network has several examples of communities where the first instance of the community did not survive the SE evaluation process and was shut down, while the second attempt resulted in a sustainable community. Focusing on attempts to establish a community on the same or similar topic with a different outcome allows us to investigate the relevance of social network structure and social trust in the sustainability of knowledge-sharing communities. They are particularly relevant if we wish to understand why some communities established themselves in their second attempt. For those pairs of communities, the topic is the same, and all the initial

[1] More information about Stack Overflow is available at: https://stackoverflow.co/ and broad introduction to Stack Exchange (SE) network is available at: https://stackexchange.com/tour. Visit https://area51.stackexchange.com/faq for more details about closed and beta SE communities and the review process.

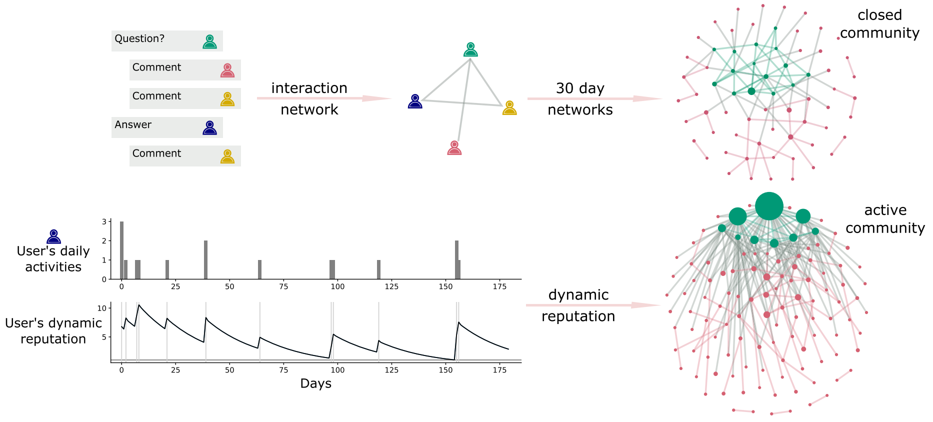

Figure 1 Visual abstract: Top row illustrates how user interaction via questions, answers, and comments is translated into an undirected network of interactions between users and finally aggregated over 30 day windows. The bottom row shows activity and corresponding dynamic reputation for one user from the closed Literature SE community. Networks on the right illustrate differences between closed and active SE Literature communities. Nodes are colored according to the core/periphery affiliation, while their size corresponds to dynamic reputation on the last day of interaction that the network contains

SE platform requirements were satisfied, but something else was crucial for community decay in the first attempt and its in the second.

Our methods and key results are summarised in a visual abstract in Fig. 1. In our main analysis, we analyze four pairs of SE communities and study the differences in the evolution of social structure and trust between closed and active communities. We have selected four topics from the STEM and humanities: astronomy, physics, economics, and literature. We focus on topics where we could find a matched pair of closed and active communities to control for the differences in topic popularity and, partially, community size. For this reason alone, we do not include Stack Overflow as the most popular community in our analysis. We analyze each pair's early stages of evolution and look at the differences between active and closed communities. Specifically, we map the interactions onto complex networks and examine how their properties evolve during the first 180 days of communities' existence. Using complex network theory [21] we quantify the structure of these networks and compare their evolution in active and closed communities on the same topic. We pay special attention to the core-periphery structure of these networks since it is one of the most prominent features of social networks [22]. We examine how core-periphery structure of active and closed communities evolve and analyze their difference. We show that active communities have a higher value of local normalized clustering and a more stable core membership. On average, the core of the sustainable communities has higher inner connectivity.

To study the evolution of social trust, we adapted the Dynamic Interaction Based Reputation Model (DIBRM) [23]. The model allows us to quantify the trust of each individual over time. We can quantify members' mean and total trust within the core and periphery and follow their evolution through time. The mean reputation of members is higher in sustainable communities than in closed ones, indicating higher levels of social trust. Furthermore, the mean reputation of core members of active communities is constantly above the mean reputation of core members in closed communities, indicating that the creation of trust in the early stages of a community's life may be crucial for its survival.

Our results show that social organization and social trust in the early phases of the life of a knowledge-sharing community play an essential role in its sustainability. Our analysis reveals differences in the evolution of these properties in communities on different topics.

The paper is organized as follows. In Sect. 2 we give a short overview of previous research. Section 3 describes the data and outlines some specific properties of each community. In Sect. 4 we describe the measures and models used for describing the local organization and measuring reputation. Section 5 shows our results. Finally, we discuss our results and selection of model parameters and time window, as well as its consequences in Sect. 6.

2 Previous research

The availability of data from the SE network led to detailed research on the different aspects of dynamics of knowledge sharing communities [6, 8–10], the roles of users [7], and their motivations to join and remain members of these communities [24–28]. The focus of the research in the previous decade was on the evolution of activity in SE communities and the different factors that influence this growth. Ahmed et al. [29] have investigated differences between technical and non-technical communities and showed that within the first four years, technical communities have a higher growth rate, more activity, and are more modular. The comparison of UX community in SE and Reddit [30] showed that the Reddit community grows faster, while SE becomes less diverse and active over time. Special attention was paid to the activities of individual users. In Ref. [31] authors argue that while the overall quality of the answers, measured in the answer score, decays over time, the quality of the answers of the individual user remains constant. This observation suggests that good answerers are *born* and not made within the community. Reputation is used as a proxy for the recognition of experts [32] by other members. However, contrary to common sense, the authors show that the presence of experts can reduce the activity of other members [32]. In [12] authors explore the role of self-and cross excitation in the temporal development of user activity. Differences between growing and declining communities and communities on STEM and humanities topics were explored. Their results show that the early stages of growing communities are characterized by the high cross-excitation of a small fraction of popular users. In contrast, later stages exhibit strong long-term self-excitation in general and cross-excitation by casual users. It was also shown that cross-excitation with power users is more important in the humanities than in STEM communities, where casual users have a more critical role.

A relatively small number of papers focus on the sustainability of SE communities. In Ref. [11], authors examine SE sites through an economic lens. They analyze the relationship between content production based on the number of participants and activities and show that an increase in the number of questions (input) increases the number of answers (output). In their works, Oliveira et al. [33] investigate activity practices and identify the tension between community spirit as proclaimed in SE guidance and individualistic values as in reputation measurement through focus groups and interviews.

Our assumption about the relevance of the structure of social networks in the sustainability of knowledge-sharing communities is supported by research on other social groups. Various factors influence the emergence [34, 35], the evolution, and the sustainability of the groups [13, 20, 36, 37]. The number of committed members [37] and the minimal level of interdependence between members [35] are important factors for the emergence of the community. The levels of activity have an important role in the emergence and stability of social groups [34, 37], while social factors, such as the size of the group, number of social contacts, or social capital, influence their emergence and collapse [13–16].

Another important branch of research of interest in the sustainability of online communities is the topic of trust. While ICTs make it easier for individuals to establish and maintain social contacts and exchange information and goods, they are also exposed to new risks and vulnerabilities. Social trust relationships, based on positive or negative subjective expectations of another person's future behavior, play an important but largely unexplored role in managing those risks. Recent works show that the vital element of trust is the notion of vulnerability in social relations, and as negative expectations of a trustee's behavior most often imply damage or harm to the trustor, decisions about which users to trust in an online community become paramount [38–40].

In communities such as SE, individuals have three sources of information to rely on when deciding to trust someone in a specific context: (1) knowledge of previous interactions, (2) expectations about future interactions, and (3) indirect information gained through a broader social network. Suppose that the number of active users in such a community increases over a more extended period. In that case, the individuals have little or no history together, no direct interactions, and almost no memory of past interactions. In that case, the social network created by the community becomes a crucial source of information. Therefore, from a network perspective, trust can be the result of reputational concerns and can flow through indirect connections linking actors to one another [40, 41].

In that case, users rely on reputation as a public measure of the reliability of other users active within the same community. Reputation is often quantified based on the history of behavior valued or promoted by a set of community norms and, as such, represents a social resource within the community [42–44]. Since reputation is public information, it is also an incentive. Agents with high reputations are motivated to act trustworthy in the future in order to preserve their status in the community [41]. This idea is supported by psychological findings suggesting that trust is primarily motivated by effects produced by the act of trust itself, regardless of more rational or instrumental outcomes of trustworthy behavior [39].

In terms of modeling collective trust and reputation in online communities, knowledge about past behaviors can be implemented in a trust model in different ways. When estimating trust between agents in a social network, graph-based models focus on the topological information, position, and centrality of agents in a social network to estimate both dyadic and collective measures of social trust. On the other hand, interaction-based models, such as the dynamic reputation model implemented in this paper (DIBRM) [23] estimate trust or reputation based on the frequency and type of agent's interactions over time without taking into account the structure and topology of the interactions between different agents in a network.

3 Data

In our main analysis, we focus on pairs of closed and active SE communities matched by topic. Astronomy, Literature, and Economics are currently active communities. All three communities thrived the second time they were proposed. The first attempt to create communities on these topics resulted in website closure within a year. We add to the comparison the early days of the Physics community and compare its evolution with the closed

Theoretical Physics community. The topics of these communities are not identical, but it is safe to assume that there is a high overlap in user demographics and interests. For these reasons, we treat this pair in the same manner as others. Furthermore, to further solidify our results we have examined the early evolution of four additional communities: Mathematics, Mathematica, Startup Business, Startups. These communities are used to inspect the robustness of our main analysis by comparing main communities with others of similar size, user growth, and activity trends.

The SE data are publicly available and released at regular time intervals. We are primarily interested in the activity and interaction data, which means that we extract the following information for posts (questions and answers) and comments: (1) for each post or comment, we extract its unique ID, the time of its creation, and unique ID of its creator - user; (2) for every question, we extract information about IDs of all answers to that question and ID of the accepted answer; (3) for each post, we collect information about IDs of its related comments. The data contains information about the official SE reputation of each user but only as a single value measuring the final reputation of the user on a day when the data archive was released. Due to this significant shortcoming, we do not include this information in our analysis. In SE, users can give positive or negative votes to questions and answers and mark questions as favorites. However, the data is again provided as a final score recorded at the release. Since this does not allow us to analyze the evolution of scores, we omit this data from our analysis.

All SE communities follow the same path from their creation until they are considered mature enough or closed. In a *Definition* phase, a small number of SE users start by designing a community by proposing hypothetical questions about a certain topic. A successful *Definition* phase is followed by a *Commitment* phase. In this phase, interested users commit to the community to make it more active. The *Beta* phase, which follows after the *Commitment* phase, is the most important. It consists of two steps: a three-week private beta phase, where only committed users may ask/answer/comment questions, and a public beta phase when other members are allowed to join the community. The duration of the public beta phase is not limited. Depending on this analysis, there are three possible outcomes: (1) the community is considered successful and it graduates; (2) the community is active but needs more work to graduate, which means that the public beta phase continues; (3) the community dies and the site is closed. The community evaluation/review process is guided by simple metrics: the average number of questions per day, average number of answers per question, percentage of answered questions, total number of users and number of avid users, and average number of visits per day. However, it should be noted that process is not straightforward and that decision criteria have substantially changed in previous years and sometimes exceptions are made for specific communities.[2]

We study how the social network properties of these social communities and the social trust created among their members evolve during the first 180 days. The first 90 days are recognized as the minimal time a newly established community should spend in the beta phase. We investigate a period that is twice as long since closed communities were active between 180 and 210 days. Given that differences in the first few months of the life of the

[2] For example, in 2022 59 websites graduated according to new criteria established in 2019 (which excluded questions per day metric), but as explained in the announcement (https://meta.stackexchange.com/questions/374096/congratulations-to-the-59-sites-that-just-left-beta) exception was made for the AI community which graduated although it didn't meet the criteria that minimum 70% questions have at least one upvoted answer.

Table 1 Community overview for first 180 days according to SE evaluation criteria

Site	Status	Answered	Questions per day	Answer ratio
Physics	Closed	83%	1.93	1.64
	Active	**93**%	**11.76**	**2.74**
Literature	Closed	79%	1.77	1.65
	Active	74%	5.04	1.10
Astronomy	Closed	**95**%	2.62	2.02
	Active	**96**%	3.57	1.49
Economics	Closed	68%	2.04	1.25
	Active	84%	5.66	1.37
Stack Exchange criteria	Excellent	>90%	>10	>2.5
	Needs some work	<80%	< 5	<1

online community can help predict its survival and evolution [45], we focus on the early evolution of SE sites.

Although the official review of SE communities in the beta phase is mostly based on simple activity indicators such as the number of questions or ratio of answers to questions,[3] these simple metrics do not provide enough information to differentiate between closed communities and those that have been proven to be sustainable in the long term. This may explain why the official guidelines for SE community review have changed and have been applied inconsistently.

Table 1 shows the values of some of these measures at 180 days point for considered communities. Although the Physics community had better metrics than Theoretical Physics and other considered communities, we see that these differences are not as apparent if we compare the remaining three pairs of communities. For instance, some of the parameters for the closed Astronomy community, for example, the percentage of answered questions and answer ratio, were better than for the community that is still active.

Another simple indicator can be the time series of active questions for the 7 days shown in Fig. 2. The question is considered active if it had at least one activity, posted answer, or comment, during the previous 7 days. The four pairs of compared communities show that active communities have a higher number of active questions after 180 days. Although this difference is evident for the Physics and Economics community, Fig. 2 shows that its value is smaller for Astronomy and Literature. Furthermore, in the case of Astronomy, the closed community had a higher number of active questions in the first 75 days.

The values of the measures shown in Tables 1 and A1 in Additional file 1, and Fig. 2 suggest that these simple measures are not good indicators of long-term sustainability. Therefore, we need a deeper understanding of the structure and dynamics of the community to understand the factors behind its sustainability. All communities must start with the same number of interesting questions, the same number of committed users, and satisfy the same thresholds to enter the public beta phase. These basic aggregated statistics are not enough to differentiate between active and closed communities. Hence, other factors determine the sustainability of communities. We investigate the role of social interaction structure and the dynamics of collective trust in the sustainability of SE communities.

[3] https://stackoverflow.blog/2011/07/27/does-this-site-have-a-chance-of-succeeding/

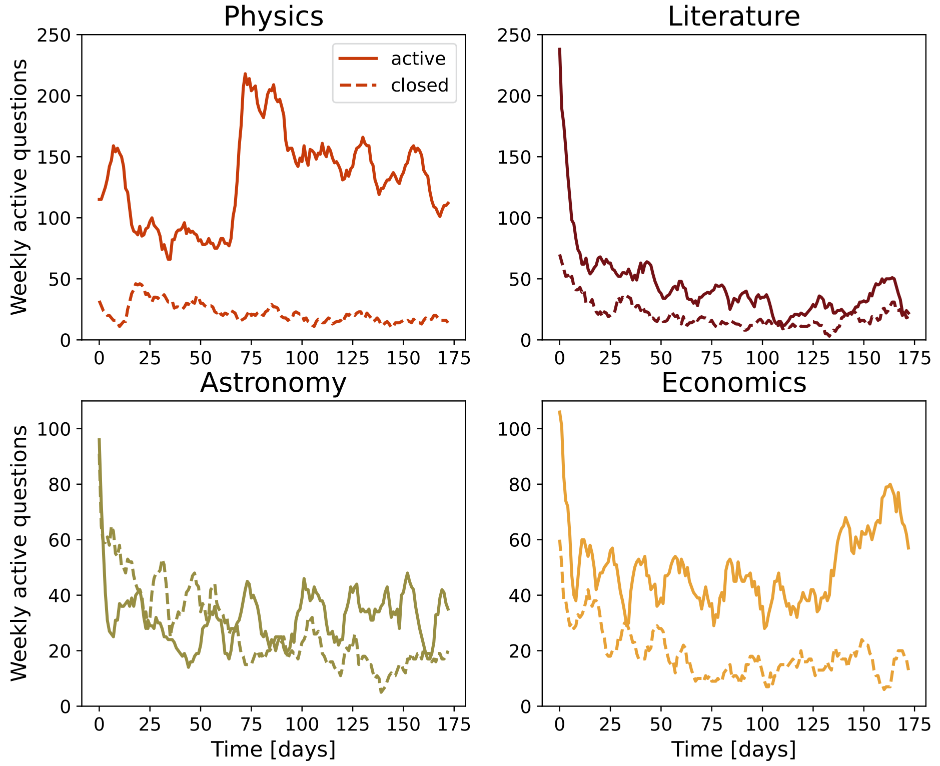

Figure 2 Variations in the number of active questions in SE communities. Number of active questions within 7 days sliding windows on the four pairs of Stack Exchange websites: Astronomy, Literature, Economics and Physics. Solid lines – active sites; dashed lines – closed sites

4 Method

We are interested in the position of trustworthy members in SE communities and how active and closed communities differ regarding this factor. First, we map the interaction data onto networks and analyze their properties and how they evolve during the first 180 days. Furthermore, we use the dynamical reputation model to estimate the trustworthiness of each member of the community and the dynamics of collective trust by studying the evolution of the mean value of reputation in the community. The entire analysis was done in Python, and the entire code for reproducing the results and figures is publicly available in an online repository.[4]

4.1 Network mapping

We treat all user interactions, answering questions, posting questions or comments, and accepting answers equally. We construct a network of users where the link between two nodes, users i and j, exists if i answers or comments on the question posted by j and vice versa, or i comments on the answer posted by j and vice versa, i accepts the answer posted by user j. We do not consider the direction or frequency of the interaction between users i and j; thus, the obtained networks are unweighted and undirected.

We create a network snapshot $G(t, t+\tau)$ at the time t for the time window length τ. Two users (i, j) are connected in a network snapshot $G(t, t+\tau)$ if they have had at least one

[4] https://github.com/ana-vranic/Stack-Exchange-communities

interaction during the time $[t, t + \tau]$. Our first network accounts for interaction within the first 30 days $G[0, 30)$, and we slide the interaction window by one day and finish with $G[149, 179)$ network. This way, we create 150 interaction networks for each community. By sliding the time window by one day, we create two consecutive networks that overlap significantly. In this way, we can capture subtle structural changes resulting from daily added/removed interactions. We calculate the different structural properties of these networks and analyze how they change over 180 days.

4.2 Clustering

There are many local and global measures of network properties [21]. These measures are not independent. However, it was shown that the degree distribution, degree-degree correlations, and clustering coefficient are sufficient to fully describe most complex networks, including social networks [46]. Furthermore, research on the dynamics of social group growth shows that links between persons' friends who are members of a social group increase the probability that that person will join that social group [47]. Successful social diffusion typically occurs in networks with a high value of the clustering coefficient [48]. These results suggest that higher local cohesion should be a characteristic of sustainable communities.

The clustering coefficient of a node quantifies the average connectivity between its neighbors and the cohesion of its neighborhood [21]. It is a probability that two neighbours of a node i are also neighbours, and is calculated using the following formula:

$$c_i = \frac{e_i}{\frac{1}{2}k_i(k_i - 1)} . \tag{1}$$

Here e_i is the number of links between the neighbours of the node i, while $\frac{1}{2}k_i(k_i - 1)$ is the maximum possible number of links determined by the degree of the node k_i. The clustering coefficient of the network C is the value of the clustering averaged over all nodes. We investigate how the clustering coefficient in an SE community changes over time by calculating its value for all network snapshots. We normalize the clustering coefficients with the value of expected clustering for the random Erdos-Renyi network with the same number of nodes N and links L: $c_{er} = p = \frac{2L}{(N(N-1))}$ [21, 49]. We compare normalized clustering coefficient for active and closed communities on the same topic to better understand the evolution of cohesion of these communities.

4.3 Core-periphery structure

Real networks, including social networks, have a distinct mesoscopic structure [22, 50]. The mesoscopic structure is manifested either through the community structure or the core-periphery structure. Networks with a community structure consist of a certain number of groups of nodes that are densely connected, with sparse connections between groups. Networks with core-periphery structures consist of two groups of nodes, with higher edge density within one group, core, and between groups. However, low edge density in the second group, periphery [22]. Research on user interaction dynamics in SE communities shows that there is a small group of highly active members who have frequent interactions with casual or low active members [8, 12]. These results indicate that we should expect a core-periphery structure in SE communities. The classification of nodes into one of these two groups provides information on their functional and dynamic roles in the network.

To investigate the core-periphery structure of SE communities and how it evolves over time, we analyze the core-periphery structure of every network snapshot. For this purpose, we use the Stochastic Block Model (SBM) adapted for the inference of the core-periphery of the network structure [22].

SBM is a model where each node belongs to one group in the given network G. For the core-periphery structure, the number of blocks is two. Thus, the elements of the vector θ_i are 1 if the node i belongs to the core or 2 for the periphery. The block connectivity matrix $\{\boldsymbol{p}\}_{2x2}$ specifies the probability p_{rs} that nodes from group r are connected to nodes in group s, where $r, s \in \{1, 2\}$.

The SBM model seeks the most probable model that can reproduce a given network G. The probability of having model parameters θ, $\boldsymbol{p}$ given network G is proportional to the likelihood of generating network G, $P(G|\theta, \boldsymbol{p})$, prior on SBM matrix $P(\boldsymbol{p})$ and prior on block assignments $P(\theta)$:

$$P(\theta, \boldsymbol{p}|G) = P(G|\boldsymbol{p}, \theta)P(\boldsymbol{p})P(\theta), \tag{2}$$

The likelihood of generating a network G is defined as:

$$P(G|\theta, \boldsymbol{p}) = \prod_{i<j} p_{r_i s_j}^{A_{ij}} (1 - p_{r_i s_j})^{1-A_{ij}}, \tag{3}$$

where the adjacency matrix element A_{ij} is equal to 1 whenever nodes i and j are connected and it is 0 otherwise.

Prior on $\boldsymbol{p}$ is the uniform distribution over all block matrices whose elements satisfy the constraint for the core-periphery structure $0 < p_{22} < p_{12} < p11 < 1$. Prior on θ consists of three parts: the probability of having 2 blocks; given the number of blocks, probability $P(n|2)$ of having groups of sizes $\{n_1, n_2\}$ and probability $P(\theta|n)$ of having particular assignments of nodes to blocks.

To fit the model, we follow the procedure set by the authors of Ref. [22] and use the Metropolis-within-Gibbs algorithm. For each 30 days snapshot network, we run 50 iterations and choose the model parameters θ and p according to the minimum description length (MDL). MDL does not change much among inferred core-periphery structures, see Fig. A1 in Additional file 1, while looking into the Adjusted Rand Index (ARI), we can notice that difference exists. Still, the ARI between pair-wise compared partitions is significant (ARI > 0.9), indicating the stability of the inferred structures. The definition and detailed descriptions of MDL and ARI are given in the Additional file 1.

4.4 Dynamic reputation model

Any dynamical trust or reputation model has to take into account distinct social and psychological attributes of these phenomena in order to estimate the value of any given trust metric [43]. First, the dynamics of trust are asymmetric, meaning that trust is easier to lose than to gain. As part of asymmetric dynamics, to make trust easier to lose, the trust metric has to be sensitive to new experiences, recent activity, or the absence of the user's activity while still maintaining the non-trivial influence of old behavior. The impact of new experiences must be independent of the total number of recorded or accumulated past interactions, making high levels of trust easy to lose. Finally, the trust metric must detect and penalize behavior that deviates from community norms.

We estimate the dynamic reputation of SE users using the Dynamic Interaction Based Reputation Model (DIBRM) [23]. This model is based on the idea of dynamic reputation, which means that the reputation of users within the community changes continuously over time: it should rapidly decrease when there is no registered activity from the specific user in the community, reputation decay, and it should grow when frequent, constant interactions and contributions to the community are detected. The highest growth in users' reputations is found through bursts of activity followed by a short period of inactivity.

Our model implementation does not distinguish between positive and negative interactions in SE communities. Therefore, we treat any interaction in the community, posting a question, answer, or comment, as a potentially valuable contribution. The evaluation criteria for SE websites that go through beta testing described in Additional file 1 do not distinguish between positive and negative interactions. The percentage of negative interactions in the communities we investigated was below 5%, see Table A2 in Additional file 1. Filtering positive interactions would also require filtering out comments because the community does not rate them. That would eliminate a large portion of direct interactions between community users, which is essential for estimating their reputation. The only negative aspect of behavior in our model is the absence of valuable contributions - the user's inactivity. This behavior can be seen as a deviation from community norms as we look at new communities in the early stages of development, where constant contributions are crucial to community growth and survival.

In DIBRM, the reputation value for each user of the community is estimated by combining two different factors: (1) *reputation growth* - the cumulative factor that represents the importance of users' activities; (2) *reputation decay* - the forgetting factor that represents the continuous decrease in reputation due to inactivity. In the case of SE communities, the forgetting factor has a literal meaning, as we can assume that active users forget users' past contributions as their attention is captured by more recent content.

In the bottom left part of Fig. 1 we see an example of reputation dynamics for a single user. There are bursts of reputation growth after multiple interactions are recorded, like in the case of two interactions in a single day recorded between days 25 and 50, followed by a period of inactivity which leads to reputation decay. In this case, the decay is interrupted by a single recorded activity before the 75th day, but then an even longer inactivity period ensued, leading to a decay that reduced the reputation of the user nearly to 0 before the 100th day. Two contrasting examples of real user reputation are explained in the Additional file 1 (Fig. A2).

Reputation dynamics revolves around the varying influence of past and recent behavior. Thus, DIBRM has two components: *cumulative factor* - estimating the contribution of the most recent activities to the overall reputation of the user; *forgetting factor* - estimating the weight of past behavior. Estimating the value of recent behavior starts with the definition of the parameter storing the basic value of a single interaction I_{b_n}. The cumulative factor I_{c_n} then captures the additive effect of successive recent interactions. In Fig. 1 we see this cumulative effect with two consecutive interactions (gray vertical lines) after day 150 which sudden jump in reputation previously reduced to zero. The reputational contribution I_n of the most recent interaction n of any given user is estimated in the following way:

$$I_n = I_{b_n} + I_{c_n} = I_{b_n}\left(1 + \alpha\left(1 - \frac{1}{S_n + 1}\right)\right) . \tag{4}$$

Here, α is the weight of the cumulative part, and S_n is the number of sequential activities. If there is no interaction at t_n, this part of interactions has a value of 0. An essential property of this component of dynamic reputation is the notion of sequential activities. Two subsequent interactions by a user are considered sequential if the time between these two activities is less than or equal to the time parameter t_a that represents the time window of interaction. This time window represents the maximum time spent by the user to make a meaningful contribution, post a question or answer, or leave a comment,

$$\Delta_n = \frac{t_n - t_{n-1}}{t_a} . \tag{5}$$

If $\Delta_n < 1$, the number of sequential activities S_n will increase by one, which means that the user continues to communicate frequently. However, large values Δ_n significantly increase the effect of the forgetting factor. This factor plays a vital role in updating the total dynamic reputation of a user at each time step, after every recorded interaction:

$$T_n = T_{n-1}\beta^{\Delta_n} + I_n . \tag{6}$$

Here, β is the forgetting factor. In our model implementation, the trust is updated each day for every user regardless of their activity status. Therefore, the decay itself is a combination of β and Δ_n: the more days pass without recorded interaction from a specific user, the more their reputation decays. Lower values of β lead to faster trust decay, as shown in Fig. A2 in the Additional file 1. In Fig. 1 we observe this long-tailed reputation loss when the user has more than 25 inactive days between days 120 and 150, reducing the reputation almost to 0.

For this work, we select the following values of these parameters: (1) we set the basic reputation contribution $I_{bn} = 1$, which means that each activity contributes 1 to the dynamical reputation; (2) for the cumulative factor α we choose the value 2 and place higher weight on recent successive interactions; (3) forgetting factor β we select the value 0.96; 4) the value of $t_a = 2$. By setting $\alpha > 1$ we enable faster growth of reputation due to a large number of subsequent interactions; see Fig. A2 in Additional file 1. Furthermore, by setting the value of $\beta < 1.0$, we increase the penalty for long inactivity periods; see Fig. A2 in Additional file 1. We discuss the selection of model parameters and their consequences in detail below. The selected values of parameters are used to measure the dynamical reputation of users in all four pair SE communities. Given these parameter values, the minimal reputation of the user immediately after having made an interaction in the SE community is 1. This reputation will decay below 1 if the user does not perform another interaction within the one-day window. Users with a reputation below the value 1 are considered inactive and *invisible* in the community; that is, their past contributions at that time are unlikely to impact other users.

4.4.1 The choice of model parameters

In this work, we used snapshots of the network of 30 days. This period corresponds to the average month, and it is common in the analyses of the structure and dynamics of social networks [51–53]. Still, there is no well-specified procedure to choose the time window. Previous studies have shown that if τ is small, subnetworks become sparse, while for too large sliding windows, some important structural changes cannot be observed [52, 54]. Thus, we have analysed how the time window choice influences our results. Figure A11 in Additional file 1 shows how considered network properties and dynamical reputation depend on the time window size for active and closed communities in case of Astronomy communities. We observe that fluctuations of all measures are more pronounced for a time window of 10 days than for 30 and 60 days. However, we find that while the structural properties of networks evolve at different rates over varied time windows, the trends remain very similar. The qualitative difference observed between closed and active communities is independent of the time window size, especially when comparing the 30 and 60 day windows. The 30-day time window ensures enough interaction, even for closed communities, while the number of observation points remains relatively high. For these reasons, we choose a sliding window of 30 days.

The initial purpose of DIBRM was to replicate the dynamics of the official SE reputation metric [23, 55]. In previous studies [55] the official SE reputation is obtained with $t_a = 2$, $\alpha = 1.4$, $\beta = 1$. This configuration of model parameters implies that there is no reputation decay and points toward the fact that the official SE reputation is hard to lose. Our application is oriented towards estimating a reputation metric which takes into account the fundamental properties of social trust, i.e. reputation decreases with members' inactivity, so we opted for a different set of parameter values.

For the basic reputation contribution of a single interaction, we selected $I_{bn} = 1$, and, at the same time, this is the threshold value of an active user. This value is intuitive as every interaction has the initial contribution of +1 to the user's reputation, although the previous works have used values of +2 and +4. Following the previous work and after examining the median/average time between subsequent interactions of the same user, we selected $t_a = 1$, which also means that the reputation in our model will be updated every day during the time window of the analysis, regardless of whether the user is active or not.

The combination of parameters α and β can significantly influence the dynamic of the single user reputation, as shown in Fig A2. We show that higher values for parameter $\alpha = 2$, highlight the burst of user activity and frequent interaction. On the other hand, the parameter beta is the forgetting factor, which at the same time determines the weight of past interactions and the reputational punishment due to user inactivity. Here, we need to select the parameter β value, so we include forgetting due to inactivity but do not penalize it too much. In Fig. A2, we show how different values of parameter β influence the time needed for a user's reputation to fall on value $I_n = 1$ due to the user's inactivity and value of dynamical reputation at the moment of the last activity. The higher the value of the parameter β and the initial dynamical reputation of the users, the longer it takes for the user's reputation to fall to the baseline value. For parameters $\beta = 0.9$ and $I_n = 5$, the user's reputation drops to value $I_n = 1$ after less than 20 days, while this time is doubled for $\beta = 0.96$. We see that for higher values of the parameter β, the time it takes for I_n to drop to 1 becomes longer and that the initial value of the reputation becomes less important.

We estimated the difference between the number of users who had at least one activity in the 30-day window and the number of users with a reputation greater than 1 during the same period for different parameter β values. We calculated the root mean square error (RMSE) between the time series of the number of active users for $\tau = 30$ and different values of β parameters; see Fig. A12 in Additional file 1. The minimal difference between these two variables is for β between 0.94 and 0.96 for both active and closed communities. Since we want to compare communities, we select $\beta = 0.96$. Our analysis reveals that the reputational decay parameter β set at 0.96 does not reduce the number of active users (based on their dynamic reputation) below the actual number of users who have been active (interacted with the community) in the time window of 30 days; see Fig. A13 in Additional file 1. Furthermore, we examine and compare the trends of two types of time series: (1) time series of active users, according to dynamical reputation; (2) time series of permanent users, users who were active in a given sliding window and continued to be active in the next one. Figure A14 in Additional file 1 shows that while the absolute number of users differs in these time series, they follow similar trends for all communities.

5 Results

5.1 Clustering and core-periphery structure of knowledge-sharing networks

We first analyze the structural properties of SE communities and examine the difference between active and closed ones. We calculate the normalized mean clustering coefficient for 30-day window networks and examine how it changes over time. Figure 3 shows the evolution of the normalized mean clustering coefficient for the eight communities. All communities that are still active are clustered, with the value of normalized clustering coefficient above 5, with Physics, the only launched community, having the highest value of normalized clustering coefficient during the first 180 days. During the larger part of the observed period, an active community's normalized clustering coefficient is higher than the normalized clustering coefficient of its closed pair. For pairs where active communities are still in the beta phase, some of closed communities have a higher value of the normalized clustering coefficient in the first 50 days. After this period, active communities have higher values of the normalized clustering coefficient. These results suggest that all communities have relatively high local cohesiveness compared to random graphs, however, the value of normalized clustering below the value 5 in the later phase of community life may indicate its decline.

Furthermore, we examine the core-periphery structure of these communities and their evolution. Specifically, we are interested in the evolution of connectivity in the core. Figure 4 shows the change in the number of links between nodes, averaged on the core nodes, $\frac{L_c}{N_c}$ over time. $\frac{2L_c}{N_c}$ is the average degree of the node in the core and, thus, $\frac{L_c}{N_c}$ is the half of the average degree. Again, the Physics community has a much higher value of this quantity than Theoretical Physics during the observed period, indicating higher connectivity between core members. Higher connectivity between core members in the active community is also characteristic of Literature. However, this quantity has the same value for active and closed communities at the end of the observation period. The differences between active and closed communities are not that prominent for Economics and Astronomy, see Fig. 4. Active and closed Economics communities have similar connectivity in the core during the first 50 days. After this period, the connectivity in the core of the active community is twice as large as in the closed community, and the difference grows at

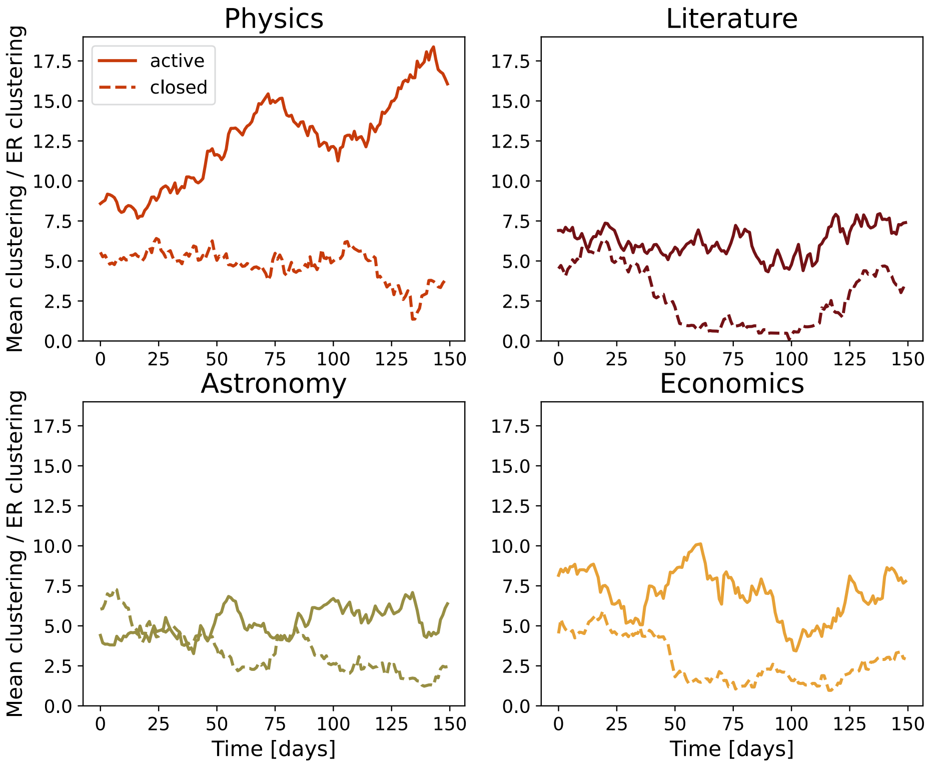

Figure 3 Normalized mean clustering coefficient of 30 days sub-networks for four pairs of Stack Exchange websites: Astronomy, Literature, Economics, and Physics. Solid lines – active sites; dashed lines – closed sites

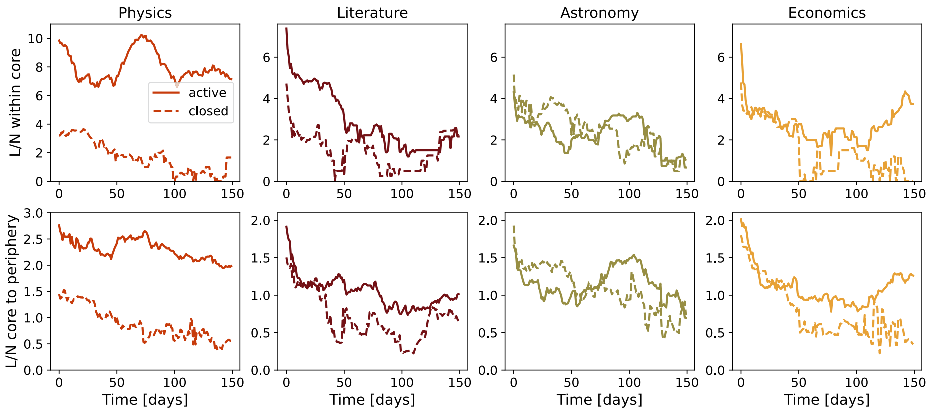

Figure 4 Connectivity among users within the core and between core and periphery. Links per node in core - top panel and links per node between core and periphery - bottom panel for the four pairs of Stack Exchange websites: Astronomy, Literature, Economics, and Physics. Solid lines – active sites; dashed lines – closed sites

the end of the observation period. The connectivity in the core of the closed Astronomy community is higher than the connectivity in the core of the active community during the first 50 days. However, as time progresses, this difference changes in favor of the active community, while this difference disappears at the end of the observation period.

The difference between active and closed communities is observed compared to the average number of core-periphery edges per network node. The connectivity between core and periphery is higher for the active communities than for the closed ones, see Fig. 4, which is very obvious if we compare Physics and Theoretical Physics communities. Moreover, the Physics community has the highest connectivity compared to all other communities. Active Literature and Economics communities have the same core-periphery connectivity as their closed counterpart. The core of the active Astronomy community has weaker connections with the periphery than the closed community during the first 50 days, see Fig. 4.

Our motivation to examine the core-periphery structure comes from reference [12]. The authors have selected 10% of the most active users and examined their mutual connectivity and connectivity with the remaining users. The split of 10% to 90% users according to their activity may appear arbitrary. The core-periphery provides a more consistent network division based on its structure. However, the connectivity patterns between popular-popular and popular-casual users, shown in Fig. A3 in Additional file 1, are similar to one observed for core-periphery in Fig. 4.

On average, the cores of active communities have a higher number of nodes than closed communities. However, the size of the core relative to the size of the network is similar for active and closed communities (Fig. A4 in Additional file 1). The size of the core fluctuates over time for active and closed communities. The core membership also changes over time. This core membership is changing more for the closed communities. We quantify this by calculating the Jaccard index between the cores of the subnetworks at the moment t_i and t_j. Figure A5 in Additional file 1 shows the value of the Jaccard index between any pair of the 150 subnetworks. The highest value of the Jaccard index is around the diagonal and has a value close to 1. The compared subnetworks are for consecutive days and have a similar structure. The value of the Jaccard index decreases with the number of days between two subnetworks $|t_i - t_j|$ faster in closed communities; see Fig. A6 in Additional file 1. This difference is the most prominent for the Literature communities, while this difference is practically non-existent for Astronomy. The relatively high value of overlap between cores of distant subnetworks for active communities further confirms that the core is more stable in these communities that in their closed counterparts.

5.2 Dynamic reputation of users within the network of interactions

To explore the differences between active and closed communities, we focus on dynamical reputation, our proxy for collective trust in these communities. The number of active users (top panel) and the mean user reputation (bottom panel) for different SE communities are shown in Fig. 5. Except in the case of Astronomy, closed communities generated less engaged users from the start and the number of active users saturated at lower values. In the case of Astronomy, the closed community started with a faster-increasing number of active users. However, within the first two months, their number dropped, while the second time around, the community started slower but kept engaging more users. Only in the still active Physics community is the number of active users an increasing function over the whole 180 day period we have observed. Panels in the bottom show mean reputation among active users, and we see that most of the time, it was higher in the still active communities than in the closed ones. The Physics community kept these mean values more stable at higher levels, whereas in other communities, we note that the initial high mean reputation decays faster. Astronomy is an exciting exception again, where we see a second sudden increase in mean user reputation, which signals an increase in user activity.

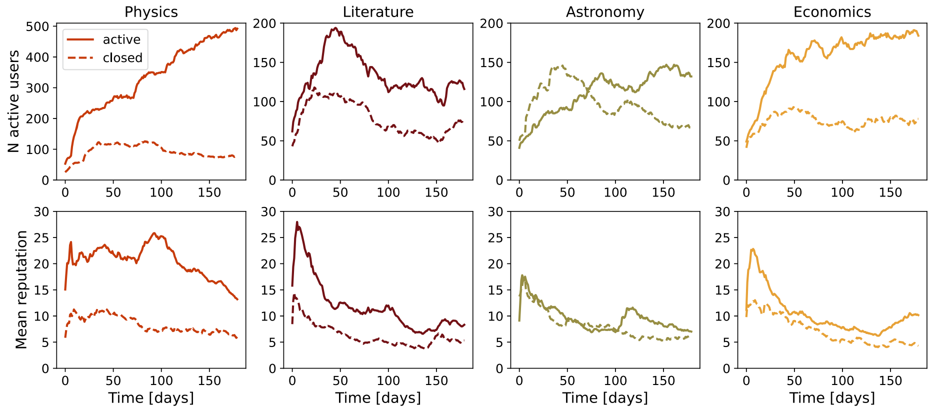

Figure 5 Active users within SE communities and their mean dynamic reputation. The number of active users (users with a reputation higher than 1) - top panel, and mean Dynamic Reputation within active users – bottom panel for the four pairs of Stack Exchange websites: Astronomy, Literature, Economics, and Physics. Solid lines – active sites; dashed lines - closed sites

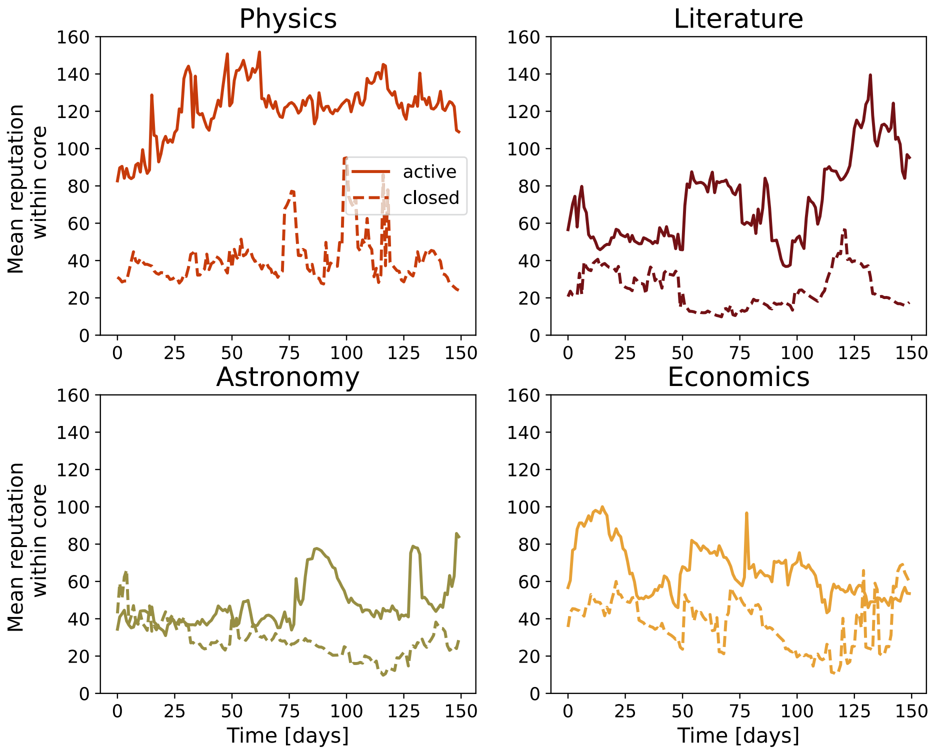

Figure 6 Mean Dynamical reputation within the core for four pairs of Stack Exchange websites: Astronomy, Literature, Economics, and Physics. Solid lines – active sites; dashed lines – closed sites

In addition, we investigate whether and how the core-periphery structure is related to collective trust in the network. Figure 6 shows the mean dynamical reputation in the core of active and closed communities and its evolution during the observation period. There are apparent differences between active and closed communities regarding dynamical reputation. The mean dynamical reputation of core users is always higher in active communities than in closed. The most significant difference is observed between the Physics and Theoretical Physics communities. The difference between active communities, which are still in the beta phase, and their closed counterparts is not as prominent. However, the active communities have a higher mean dynamical reputation, especially in the later phase of the observation period. The only difference in the pattern is observed for Astronomy communities at the early stage of their life. The closed community has a higher value of dynamic reputation than the active community. This observation is in line with similar patterns in the evolution of mean clustering, core-periphery structure, and mean reputation.

By definition, the core consists of very active individuals. Thus we expect a higher total dynamical reputation of users in the core than the total reputation of users belonging to subnetworks periphery. Figure A7 in Additional file 1 shows the ratio between the total reputation of the core and periphery for closed and active communities and their evolution. The ratio between the total reputation of core and periphery in Physics is always higher than in the Theoretical Physics community. A similar pattern can be observed for Literature communities, although the difference is not as prominent as in the case of Physics. The ratio of total dynamical reputation between core and periphery was higher in the closed Economics community during the early days of its existence. However, this ratio becomes higher for active communities in the later stage of their lives. Communities around the astronomy topic deviate from this pattern, which shows the specificity of these two communities.

To complete the description of the evolution of dynamic reputation, we examine the evolution of the Gini index of dynamical reputation among the active members of SE sites, shown in Fig. A8 in Additional file 1. Both closed and active communities have high values of the Gini index, indicating that the dynamic reputation is distributed unequally among users. Notably, all communities have the highest Gini index at the start, signaling that the inequality in users' activity at the start, and thus their dynamic reputation is the highest. After this initial peak, the Gini index decreases, but it persists at higher levels in communities that are still active than in the closed ones, except in the case of the Astronomy community. In this case, the active community had a higher Gini index until just before the observation period, when the Gini coefficient increased in the closed community.

Figure A9 in Additional file 1 shows the evolution of the assortativity coefficient for users' dynamical reputation. The observed networks are disassortative during the most significant part of 180 days period. Users with high dynamical reputations tend to connect with users with a low value of dynamical reputation in all eight communities. We also compare the degree and betweenness centrality of the users and their dynamical reputation by calculating the correlation coefficient between these measures for each sliding window, see Fig. A10 and detailed explanation in Additional file 1. The correlation between these centrality measures and dynamical reputation is very high. In active communities on physics, economics, and literature topics, the correlation between centrality measures and users' reputation is exceptionally high, above 0.85, and does not fluctuate much during the observation period. There is a clear difference between active and closed communities for these three pairs. The Astronomy pair deviates from this pattern for the first 100 days. After this period, the pattern is similar to one observed for the other three pairs of communities. The results reveal that degree and betweenness centrality are correlated more with a reputation in active than in closed communities.

6 Discussion and conclusions

In this work, we have explored whether the structure and dynamics of social interactions determine the sustainability of knowledge-sharing communities. We have adopted a model of dynamical reputation to measure the collective trust of members and analyzed its dynamics. For this purpose, we use the data from the SE platform of knowledge-sharing communities where members ask and answer questions on focused topics. We selected four pairs of active and closed communities on the same or similar topic. Specifically, two topics are from the STEM field, physics, and astronomy, and two are from social sciences and humanities, economics and literature.

We have examined the evolution of the normalized average clustering coefficient in closed and active SE communities. Our results show that active communities have significantly higher values of clustering coefficient compared to ER graphs of the same size in the later phase of community life than closed communities. In the early phase of communities' lives, the clear difference between active and closed communities is observed only for the physics topic; see Fig. 3. The high value of the normalized clustering coefficient observed for the active Physics community suggests that communities with high local cohesiveness are sustainable and mature faster than others.

The core in active communities is more strongly connected with the periphery than in closed communities, indicating that active members engage more often with occasionally active members; see Fig. 4. These results suggest that active communities are more inclusive than closed ones. Furthermore, our analysis shows that average connectivity between core members is not as crucial to community sustainability as expected. Although active Physics and Economics communities exhibit much higher connectivity in the core than their closed counterparts, this is not true for communities focused on astronomy and literature. However, our results show that a member's lifetime in the core is longer for active communities, indicating a more stable core in active communities.

Analysis of the evolution of the core-periphery and its connectivity patterns suggests a higher trust between active and sporadically active members. To further explore this, we have adapted the dynamical reputation model [23], which allowed us to follow the evolution of trust of each member.

The total dynamical reputation of core members during their first 180 days was higher for active communities than for their closed counterparts. While relative core size is less than 40%, Fig. A4 in Additional file 1, the ratio between the total reputation of nodes in the core and ones in the periphery is consistently above 0.5, indicating that the average reputation of members in the core is higher than the reputation of the node in the periphery. The ratio between the total reputation of core and periphery nodes has a higher value in the active community of Physics, Literature, and Economics. For most of the 180 days, this ratio has a value higher than one. The Astronomy communities are outliers, but the core members have a higher total reputation than members on the periphery, even for these two communities. Our results imply that the most trusted members in the community are the core members, who also generate more trust in active communities. They have a higher reputation generated through interactions with both core and nodes in the periphery, see Fig. 6. Furthermore, the overall levels of trust are higher in active communities, which is reflected in the fact that the mean user reputation is higher in these communities; see Fig. 5.

The choice of the topics and selection of SE communities of a various number of users, question, answer and comments, see Table A1 in the Additional file 1, guarantees, up to a certain extent, the generality of our results. However, there are certain limitations to the generalizability of our findings. While SE communities provide very detailed data that enable the study of the structure and dynamics of knowledge-sharing communities, we must not ignore the fact that they have some properties that make them specific.

SE communities are about specific topics; they mostly bring together people who are passionate about or are experts in a specific field. These communities attract people from the general population. Since we were interested in excluding the factor of the topic in our research, we studied and compared active and closed communities on the same topic. In the SE network, these pairs of communities are pretty rare, which has substantially limited our sample size, leaving the possibility for the occurrence of outliers that do not follow our general conclusions.

To further solidify our results, we have examined the early evolution of four additional communities: Mathematics, Mathematica, Startup Business, and Startups. Mathematics and Mathematica communities graduated early in the process, while both communities on startup topics were closed after spending some time in the public beta phase. Figures A15 and A16 in the Additional file 1 show that both communities on the subject of mathematics exhibit a similar evolutionary path as the Physics community. They have a high mean reputation, stable and relatively large cores with high average trustworthiness of core members, see Fig. A15 in Additional file 1. While the numbers of active users in these two communities and the Physics community differ, we see that this does not influence the average reputation of users or the size of the core. This is even more evident if we compare the Physics community with the closed Startup Business community. We see from Fig. A16 in Additional file 1 that the number of active users grows much faster for this community than for Physics. However, the average reputation in the community is comparable with the ones that were eventually closed, Theoretical Physics and Startups. Furthermore, the core size is comparable with the core of Physics, but the average trustworthiness of core members is similar to one for closed communities. These results demonstrate that even the communities with high early activity and a number of active users will not become sustainable if they do not develop a core of trustworthy members. Startups community has a behavior very similar to Theoretical Physics community. The comparison between two startup communities, shows that despite their difference in the activity levels these communities have similar evolution path during the first 180 days.

We have also decided to map interactions to networks so that the resulting network is unweighted and undirected. We use unweighted edges for a finer distinction between the structure and community dynamics. The number of repeated user interactions is captured with dynamic reputation, while the edges carry only structural information without the number of repeated interactions. Furthermore, as we map interactions to networks using sliding windows, the repeated presence of an edge throughout different windows gives us partial information about the durability and the frequency of the dyadic relationship. Similarly, we opted against directed weights as we are not interested in diffusion or flow of information and undirected edges represent a more parsimonious view of the community structure. However, these choices did have consequences in the choice of core-periphery detection method, and it is possible that with different network mapping, other methods would prove more suitable.

Finally, there are many ways to measure collective trust and reputation in online social communities. We have selected the dynamical reputation model because it was developed to measure reputation in SE communities. Furthermore, the model allowed us to study the evolution of trust in communities. However, the model requires fine-tuning of its parameters and does not distinguish positive from negative interactions. We have selected our parameters to replicate the activity of the SE communities in the time window of $\tau = 30$ days. Our analysis shows that while the choice of the sliding window, τ, may seem arbitrary, the different values do not influence the general conclusions; see Fig. A11 in Additional file 1. The interactions in SE communities are mostly not emotional, and thus, the model is suitable for measuring collective trust in these communities. However, the interaction in other knowledge-sharing communities can be much more emotional, and therefore the dynamical reputation model needs to be adapted to measure reputation in these communities.

Our results show that the trustworthiness of core members thus represents one of the essential parameters for determining community sustainability. Sustainable communities have a core of trustworthy members. The core of sustainable communities is more densely connected, and its connectivity with the periphery is more significant than in closed communities. The observed feature is especially prominent in the Physics community, which is the only active community considered to be mature. As we stated, active communities on topics of astronomy, economics and literature were in the beta phase. However, since December 2021,[5] these communities graduated. The core of sustainable communities exhibits higher degrees of stability during their first 180 days. Sustainable communities have higher local cohesiveness, which is reflected in the relatively high value of the normalized clustering coefficient. Our results show that these conclusions hold for both STEM and humanities topics. However, we do not observe apparent differences between active and closed Astronomy communities for some quantities. In the case of Astronomy and sometimes Economics, we find that closed communities had higher normalized clustering coefficients and higher core-core and core-periphery connectivity during the early phase of community life. These observations suggest that the properties of the network during the early phase of the community's existence may lead to wrong conclusions about its sustainability. Our results also imply that information about community sustainability is hidden in the evolution of different network and trust properties.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1140/epjds/s13688-023-00381-x.

Additional file 1. The file contains all additional figures, tables and descriptions regarding the analysis performed in the manuscript. The file is in pdf format. (PDF 3.6 MB)

Acknowledgements

Numerical simulations were run on the PARADOX-IV supercomputing facility at the Scientific Computing Laboratory, National Center of Excellence for the Study of Complex Systems, Institute of Physics Belgrade.

Funding

AA, AV and MMD acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science, and Technological Development of the Republic of Serbia.

[5] https://stackoverflow.blog/2021/12/16/congratulations-are-in-order-these-sites-are-leaving-beta/

Abbreviations

ARI, Adjusted Rand Index; DIBRM, Dynamic Interaction Based Reputation Model; ICT, Information and communication technologies; MDL, Minimum Description Length; RMSE, Root mean square error; SBM, Stochastic Block Model; SE, Stack Exchange.

Availability of data and materials

The Stack Exchange data can be downloaded from Stack Exchange Data Dump, https://archive.org/details/stackexchange. Area 51 Stack Exchange communities can be downloaded from https://area51.stackexchange.com/. The source code and the datasets generated and analysed during the current study are publicly available at https://github.com/ana-vranic/Stack-Exchange-communities.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author contribution

AV, AT, AA, MMD designed the research. AV, AT and AA collected the data and performed data analysis. All authors wrote and edited the final manuscript. All authors read and approved the final manuscript.

Author details

[1]Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, Belgrade, Serbia. [2]Department of Sociology, Faculty of Philosophy, University of Novi Sad, Novi Sad, Serbia. [3]Two desperados, Belgrade, Serbia.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 5 July 2022 Accepted: 14 February 2023 Published online: 24 February 2023

References

1. Leydesdorff L (2001) In: A sociological theory of communication: the self-organization of the knowledge-based society. Universal-Publishers, USA. https://doi.org/10.1108/jd.2002.58.1.106.2
2. Leydesdorff L (2012) The triple helix, quadruple helix,…, and an n-tuple of helices: explanatory models for analyzing the knowledge-based economy? J Knowl Econ 3(1):25–35. https://doi.org/10.1007/s13132-011-0049-4
3. Lipkova H, Landová H, Jarolímková A (2017) Information literacy vis-a-vis epidemic of distrust. In: European conference on information literacy. Springer, Berlin, pp 833–843
4. Lucassen T, Schraagen JM (2012) Propensity to trust and the influence of source and medium cues in credibility evaluation. J Inf Sci 38(6):566–577
5. Abrahao B, Parigi P, Gupta A, Cook KS (2017) Reputation offsets trust judgments based on social biases among airbnb users. Proc Natl Acad Sci 114(37):9848–9853
6. Dankulov MM, Melnik R, Tadić B (2015) The dynamics of meaningful social interactions and the emergence of collective knowledge. Sci Rep 5(1):1–10. https://doi.org/10.1038/srep12197
7. Saxena A, Reddy H (2021) Users roles identification on online crowdsourced q&a platforms and encyclopedias: a survey. J Comput Soc Sci 1–33. https://doi.org/10.1007/s42001-021-00125-9
8. Santos T, Walk S, Kern R, Strohmaier M, Helic D (2019) Activity archetypes in question-and-answer (q8a) websites—a study of 50 stack exchange instances. ACM Trans Soc Comput 2(1):1–23. https://doi.org/10.1145/3301612
9. Slag R, de Waard M, Bacchelli A (2015) One-day flies on stackoverflow-why the vast majority of stackoverflow users only posts once. In: 2015 IEEE/ACM 12th working conference on mining software repositories. IEEE, pp 458–461. https://doi.org/10.1109/MSR.2015.63
10. Chhabra A, Iyengar SRS (2020) Activity-selection behavior of users in stackexchange websites. In: Companion proceedings of the web conference 2020, pp 105–106. https://doi.org/10.1145/3366424.3382720
11. Dev H, Geigle C, Hu Q, Zheng J, Sundaram H (2018) The size conundrum: why online knowledge markets can fail at scale. In: Proceedings of the 2018 world wide web conference, pp 65–75. https://doi.org/10.1145/3178876.3186037
12. Santos T, Walk S, Kern R, Strohmaier M, Helic D (2019) Self-and cross-excitation in stack exchange question & answer communities. In: The world wide web conference, pp 1634–1645. https://doi.org/10.1145/3308558.3313440
13. Oliver PE, Marwell G (2001) Whatever happened to critical mass theory? A retrospective and assessment. Sociol Theory 19(3):292–311. https://doi.org/10.1111/0735-2751.00142
14. Smiljanić J, Mitrović Dankulov M (2017) Associative nature of event participation dynamics: a network theory approach. PLoS ONE 12(2):0171565. https://doi.org/10.1371/journal.pone.0171565
15. Török J, Kertész J (2017) Cascading collapse of online social networks. Sci Rep 7(1):16743. https://doi.org/10.1038/s41598-017-17135-1
16. Lőrincz L, Koltai J, Győr AF, Takács K (2019) Collapse of an online social network: burning social capital to create it? Soc Netw 57:43–53. https://doi.org/10.1016/j.socnet.2018.11.004
17. Wasko MM, Faraj S (2005) Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. MIS Q 29(1):35–57. https://doi.org/10.2307/25148667
18. Hung S-Y, Durcikova A, Lai H-M, Lin W-M (2011) The influence of intrinsic and extrinsic motivation on individuals' knowledge sharing behavior. Int J Hum-Comput Stud 69(6):415–427. https://doi.org/10.1016/j.ijhcs.2011.02.004
19. Rode H (2016) To share or not to share: the effects of extrinsic and intrinsic motivations on knowledge-sharing in enterprise social media platforms. J Inf Technol 31(2):152–165. https://doi.org/10.1057/jit.2016.8

20. Kairam SR, Wang DJ, Leskovec J (2012) The life and death of online groups: predicting group growth and longevity. In: Proceedings of the fifth ACM international conference on web search and data mining, pp 673–682. https://doi.org/10.1145/2124295.2124374
21. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D-U (2006) Complex networks: structure and dynamics. Phys Rep 424(4–5):175–308. https://doi.org/10.1016/j.physrep.2005.10.009
22. Gallagher RJ, Young J-G, Welles BF (2021) A clarified typology of core-periphery structure in networks. Sci Adv 7(12):9800. https://doi.org/10.1126/sciadv.abc9800
23. Melnikov A, Lee J, Rivera V, Mazzara M, Longo L (2018) Towards dynamic interaction-based reputation models. In: 2018 IEEE 32nd international conference on Advanced Information Networking and Applications (AINA), pp 422–428. https://doi.org/10.1109/AINA.2018.00070
24. Wei X, Chen W, Zhu K (2015) Motivating user contributions in online knowledge communities: virtual rewards and reputation. In: 2015 48th Hawaii international conference on system sciences. IEEE, pp 3760–3769. https://doi.org/10.1109/HICSS.2015.452
25. Yanovsky S, Hoernle N, Lev O, Gal K (2019) One size does not fit all: badge behavior in q&a sites. In: Proceedings of the 27th ACM conference on user modeling, adaptation and personalization, pp 113–120. https://doi.org/10.1145/3320435.3320438
26. Santos T, Burghardt K, Lerman K, Helic D (2020) Can badges Foster a more welcoming culture on q&a boards? In: Proceedings of the international AAAI conference on web and social media, vol 14, pp 969–973
27. Bornfeld B, Rafaeli S (2019) When interaction is valuable: feedback, churn and survival on community question and answer sites: the case of stack exchange. In: Proceedings of the 52nd Hawaii international conference on system sciences
28. Kang M (2021) Motivational affordances and survival of new askers on social q&a sites: the case of stack exchange network. Journal of the Association for Information Science and Technology. https://doi.org/10.1002/asi.24548
29. Ahmed S, Yang S, Johri A (2015) Does online q&a activity vary based on topic: a comparison of technical and non-technical stack exchange forums. In: Proceedings of the second (2015) ACM conference on learning@ scale, pp 393–398. https://doi.org/10.1145/2724660.2728701
30. Chen G, Mok L (2021) Characterizing growth and decline in online ux communities. In: Extended abstracts of the 2021 CHI conference on human factors in computing systems, pp 1–7. https://doi.org/10.1145/3411763.3451646
31. Posnett D, Warburg E, Devanbu P, Filkov V (2012) Mining stack exchange: expertise is evident from initial contributions. In: 2012 international conference on social informatics. IEEE, pp 199–204. https://doi.org/10.1109/SocialInformatics.2012.67
32. Pal A, Chang S, Konstan JA (2012) Evolution of experts in question answering communities. In: Sixth international AAAI conference on weblogs and social media
33. Oliveira N, Muller M, Andrade N, Reinecke K (2018) The exchange in stackexchange: Divergences between stack overflow and its culturally diverse participants. Proc ACM Hum-Comput Interact 2(CSCW):1–22. https://doi.org/10.1145/3274399
34. Dover Y, Kelman G (2018) Emergence of online communities: empirical evidence and theory. PLoS ONE 13(11):0205167. https://doi.org/10.1371/journal.pone.0205167
35. Han X, Cao S, Shen Z, Zhang B, Wang W-X, Cressman R, Stanley HE (2017) Emergence of communities and diversity in social networks. Proc Natl Acad Sci 114(11):2887–2891. https://doi.org/10.1073/pnas.1608164114
36. Kleineberg K-K, Boguñá M (2015) Digital ecology: coexistence and domination among interacting networks. Sci Rep 5(1):1–11. https://doi.org/10.1038/srep10268
37. Oliver P, Marwell G, Teixeira R (1985) A theory of the critical mass. I. Interdependence, group heterogeneity, and the production of collective action. Am J Sociol 91(3):522–556. https://doi.org/10.1086/228313
38. Dunning D, Anderson JE, Schlösser T, Ehlebracht D, Fetchenhauer D (2014) Trust at zero acquaintance: more a matter of respect than expectation of reward, vol 107 pp 122–141. https://doi.org/10.1037/a0036673
39. Dunning D, Fetchenhauer D, Schlösser T (2019) Why people trust: solved puzzles and open mysteries. Curr Dir Psychol Sci 28(4):366–371. https://doi.org/10.1177/0963721419838255
40. Schilke O, Reimann M, Cook KS (2021) Trust in Social Relations. Annu Rev Sociol 47(1):239–259. https://doi.org/10.1146/annurev-soc-082120-082850
41. McEvily B, Zaheer A, Soda G (2021) Network trust. In: Gillespie N, Fulmer A, Lewicki R (eds) Understanding trust in organizations. Taylor & Francis. https://doi.org/10.4324/9780429449185
42. Aberer K, Despotovic Z (2001) Managing trust in a peer-2-peer information system. In: CIKM'01. Association for Computing Machinery, New York, pp 310–317. https://doi.org/10.1145/502585.502638
43. Duma C, Shahmehri N, Caronni G (2005) Dynamic trust metrics for peer-to-peer systems. In: 16th international workshop on database and expert systems applications (DEXA'05). IEEE, pp 776–781. https://doi.org/10.1109/DEXA.2005.80
44. Tschannen-Moran M, Hoy W (2000) A multidisciplinary analysis of the nature, meaning, and measurement of trust. In: Review of educational research, vol 70. American Educational Research Association, pp 547–593. https://doi.org/10.3102/00346543070004547
45. Dover Y, Goldenberg J, Shapira D (2020) Sustainable online communities exhibit distinct hierarchical structures across scales of size. Proc R Soc A 476(2239):20190730. https://doi.org/10.1098/rspa.2019.0730
46. Orsini C, Dankulov MM, Colomer-de-Simón P, Jamakovic A, Mahadevan P, Vahdat A, Bassler KE, Toroczkai Z, Boguná M, Caldarelli G et al (2015) Quantifying randomness in real networks. Nat Commun 6(1):8627. https://doi.org/10.1038/ncomms9627
47. Backstrom L, Huttenlocher D, Kleinberg J, Lan X (2006) Group formation in large social networks: membership, growth, and evolution. In: Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining, pp 44–54. https://doi.org/10.1145/1150402.1150412
48. Centola D, Eguíluz VM, Macy MW (2007) Cascade dynamics of complex propagation. Phys A, Stat Mech Appl 374(1):449–456. https://doi.org/10.1016/j.physa.2006.06.018
49. Bollobás B, Riordan OM (2003) Mathematical results on scale-free random graphs. In: Handbook of graphs and networks: from the genome to the Internet, pp 1–34

50. Fortunato S (2010) Community detection in graphs. Phys Rep 486(3–5):75–174. https://doi.org/10.1016/j.physrep.2009.11.002
51. Saramäki J, Moro E (2015) From seconds to months: an overview of multi-scale dynamics of mobile telephone calls. Eur Phys J B 88(6):1–10. https://doi.org/10.1140/epjb/e2015-60106-6
52. Krings G, Karsai M, Bernhardsson S, Blondel VD, Saramäki J (2012) Effects of time window size and placement on the structure of an aggregated communication network. EPJ Data Sci 1(1):1. https://doi.org/10.1140/epjds4
53. Barrat A, Gelardi V, Le Bail D, Claidiere N (2021) From temporal network data to the dynamics of social relationships. Proc R Soc Lond B, Biol Sci 288:20211164. https://doi.org/10.1098/rspb.2021.1164
54. Arnold NA, Steer B, Hafnaoui I, Parada GHA, Mondragon RJ, Cuadrado F, Clegg RG (2021) Moving with the times: investigating the alt-right network gab with temporal interaction graphs. Proc ACM Hum-Comput Interact 5(CSCW2) 447. https://doi.org/10.1145/3479591
55. Yashkina E, Pinigin A, Lee J, Mazzara M, Adekotujo AS, Zubair A, Longo L (2019) Expressing trust with temporal frequency of user interaction in online communities. In: Advanced information networking and applications. Springer, Cham. https://doi.org/10.1007/978-3-030-15032-7_95