RANDOM UTILITY COORDINATION GAMES ON NETWORKS

MARCIN PĘSKI

Abstract. We study static binary coordination games with random utility played on networks. In equilibrium, each agent chooses an action only if a fraction of her neighbors with the same action is higher than a agent-specific i.i.d. threshold. A *fuzzy convention* x is a profile where (almost) all agents choose the high action if their threshold is smaller than x and low action otherwise. The *random-utility (RU) dominant outcome* x^* is a maximizer of an integral of the distribution of thresholds. The definition generalizes Harsanyi and Selten (1988)'s risk dominance to coordination games with random utility. We show that, on each sufficiently large and fine network, there is an equilibrium that is a fuzzy convention x^*. On some networks, including a city network, all equilibria are fuzzy conventions x^*. Finally, fuzzy conventions x^* are the only behavior that is robust to misspecification of the network structure.

1. Introduction

An individual's behavior in social or economic situations is often positively influenced by similar decisions made by their friends, acquaintances, or neighbors. An important recent example is the post-Covid era mask-wearing: some people wear masks to protect themselves or others, others don't wear them because of inconvenience or personal beliefs, and many, including the author of this paper, are positively affected by how many people around them wear masks. Other examples include the decision to maintain a neat front yard, to obey speed limits or tax laws, to engage in criminal activity, or to adopt a technology with network externalities. A large literature has established

Date: September 23, 2024.

I am very grateful to Stephen Morris and Philip Neary for detailed comments, Jacopo Perego for discussion, Denise Cruz for the main example, and an anonymous referee for insightful suggestions. S. Morris's comments greatly influenced the proof of Theorem 1. I gratefully acknowledge financial support from the Insight Grant of the Social Sciences and Humanities Research Council of Canada.

conditions under which a particular behavior becomes a convention: it is adopted by everyone (see Young (1993), Ellison (1993), Morris (2000), among many others). These results typically assume that agents have almost identical preferences, and show that a contagion-like process, possibly initiated by a small perturbation to the preferences, leads to uniformity.

At the same time, a completely uniform behavior is rarely observed in the real world. Even in situations which clearly involve positive externalities, there will often be interactions in which neighbors make opposite choices. An obvious reason is that individuals are different and their tastes and unique circumstances play just as important of a role in determining their decisions as the behavior of their neighbors. The goal of this paper is to study coordination games with heterogeneous payoffs with two questions in mind. First, is there a useful and coherent way in which heterogeneous-behavior equilibria can be understood as conventions? Second, can we explain how people coordinate on a convention? Are some conventions more natural than others?

For this purpose, we study a random utility binary coordination game played in a network. Each network node contains a single agent who interacts with her neighbors. We are interested in the asymptotic of equilibrium behavior as the network becomes arbitrarily large and, importantly, as the graph becomes sufficiently fine i.e., the weight of the largest neighbor in a neighborhood of each agent becomes sufficiently small. The latter ensures that no single individual has a disproportionate impact on another, and it is the first key assumption in our model.

Each agent chooses a binary (high or low) action and the relative gain from the action is increasing in the fraction of neighbors who make the same choice. Each agent has a individual threshold τ_i, with the interpretation that the high action is the agent's best response if and only if more than fraction τ_i of her neighbors do the same. Thresholds are distributed i.i.d., with distribution given by cdf $P(.)$. The independence assumption is the second key assumption of our model and it is appropriate for some but not all applications. An example of cdf $P(.)$ is drawn on Figure 1; for each x, $P(x)$ is the fraction of the population with a threshold equal to or smaller than x. Importantly, unlike in the coordination literature mentioned above, the level of preference heterogeneity captured by $P(.)$ is non-zero and and non-disappearing.

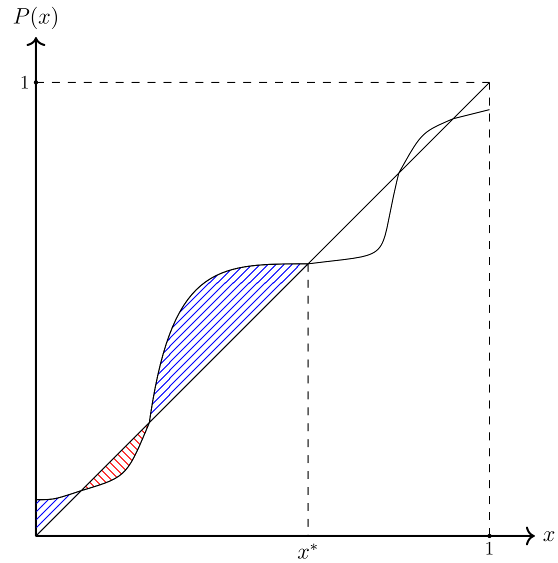


FIGURE 1. Threshold cdf P.

A conceptual contribution of this paper is a definition of a convention appropriate for large random utility coordination games. Define a fuzzy convention x as an action profile where almost all agents choose the high action if $\tau_i < x$ and the low action if $\tau_i > x$. If x is an atom of distribution $P(.)$, the definition allows for randomization at $\tau_i = x$. Our assumptions on networks imply that, in a fuzzy convention, almost all agents observe approximately $P(x)$ fraction of their neighbors choosing the high action. This definition captures individual heterogeneity of actions, with two types of uniformity: (a) almost all agents choose their action as the same function of their threshold and (b) almost all agents experience almost the same average behavior of their neighbors. For a fuzzy convention x to be an equilibrium, the choice in (a) must be a best response, which implies that it is an intersection with 45°-degree line, $x = P(x)$. Figure 1 illustrates with multiple candidate solutions.

Next, we define a particular fixed-point. Let *random utility-dominant*, or *RU-dominant*, outcome x^* be a solution to the maximization problem

$$x^* \in \arg\max_x \int_0^x \left(y - P^{-1}(y)\right) dy. \tag{1}$$

The definition implies that $P(x^*) = x^*$. Geometrically, the maximized objective on the right-hand side is equal to the area above 45 degree line and below function P (blue area on (see Figure 1) minus the area below 45 degree line and above P (red area). The RU-dominant outcome depends on the threshold distribution, and generically, it is unique. Two observations about special cases of our model motivate this definition further. First, (1) is equivalent to a formula from Morris and Shin (2006), where it is derived as a potential function for the continuum population version of the model where agents treat the entire population as their neighbors. Additionally, if the threshold distribution is concentrated on a single outcome (i.e., all agents' preferences are identical), then the RU-dominant outcome is equivalent to the standard risk-dominant outcome of a 2×2 coordination game (Harsanyi and Selten (1988)).

The results of the paper show that fuzzy convention x^* is the "right" solution: Informally, all networks have an equilibrium that is a fuzzy convention x^*, and, on some networks, there are no other equilibria. More precisely, first, we show that for each network that is sufficiently large and fine, with a probability close to 1 (i.e., for almost all realizations of thresholds), there is an equilibrium that is a fuzzy convention x^*. The proof relies on a characterization of coordination games as potential games. (For an arbitrary network, a potential function is necessarily different than the one in (1).) Such games are introduced in Monderer and Shapley (1996), where it is shown that any profile that is a local maximizer of the potential function is an equilibrium of the underlying game. In the proof, we show that, regardless of the structure of the network, with a probability close to 1, the *global* maximizer of the potential function is a fuzzy convention x^*. The difficult part of the proof is to derive a version of a uniform law of large numbers and to show that it guarantees that action profiles that are not fuzzy conventions x^* cannot maximize the potential.

Second, we show there exist networks, where, with a large probability, all equilibria are fuzzy convention x^*. An example of such a network is a city-like network, where agents live on a 2-dimensional grid lattice and they interact with agents in a sufficiently large neighborhood. The idea of the proof is to show that, for each profile with an average behavior that is not *RU*-dominant, contagion-like best response dynamics would bring the behavior close to x^*. The proof uses an idea from Blume (1995a) and Lee and Valentinyi (2000) (see also Morris (2000)) to show how a contagion wave spreads across lattice networks. There are two novel difficulties relative to earlier literature. First, unlike in the earlier literature, the agent preferences are random and heterogeneous. Instead of binary wave (where there is a sharp separation between risk-dominated and risk-dominant regions), the contagion wave here has multiple values as it describes fraction of agents that adopt the high action. Additionally, we must compare the likelihood that a favorable configuration of payoff shocks may initiate such a wave, with the likelihood that such a wave would not be stopped by an unfavorable configuration of payoff shocks. The problem with the latter is the reason why the 1-dimensional network of Ellison (1993) is not a good example for the result and a 2- (or more) dimensional lattice is needed.

The two results together suggest that RU-dominant outcome x^* is the only prediction of aggregate equilibrium behavior that is network-independent. We formalize this through a definition that is inspired by Kajii and Morris (1997): Consider an analyst who predicts agent's behavior but she is not certain whether her model correctly specifies the network interactions, or whether the agents know the entire network. We say that the behavior is robust to misspecifications if, even if she or the agents are wrong, her prediction is close to some equilibrium of the true model. Our results imply that fuzzy convention x^* is the only robust prediction.

1.1. **Literature review.** This is the first paper with predictions about behavior in static complete-information random-utility games on networks. The model and techniques used draw from two strands of the literature: random utility games on networks and models of learning (or evolution) in games.

The first random-utility coordination model was introduced in Granovetter (1978). Granovetter works with a complete (continuum) network, where the agents' payoffs depend on the average behavior in the entire population. A large literature generalized Granovetter's model to networks. Typically, each agent is a single node on a network and adopts the new behavior (for example, wears a mask) only if the fraction of her neighbors doing the same is larger than her threshold. Many papers, like Watts (2002), or López-Pintado (2008) (among many others) study Granovetter's model on a Erdos-Renyi style of a random graph with heterogeneous degree distribution. The limitation of such models is that they do not capture many important aspects of real-world networks, like clustering, or overlapping neighborhoods which are known to play important role in coordination or contagion phenomena.

Jackson and Yariv (2007) (see also Galeotti *et al.* (2010)) analyzes a Bayesian equilibrium, where the agents choose their action without knowing the thresholds of their neighbors. This assumption improves model's tractability as the agent's behavior does not depend on individual thresholds of her neighbors. At the same time, this assumption is not satisfactory if the equilibrium is to be interpreted as a long-term process as each agent may change her behavior when they observe the actions of their neighbors. This is the first key difference from our model, where an equilibrium is a steady state behavior *after* the thresholds are realized and actions are chosen. Because our model is a static, complete information equilibrium for given realization of thresholds, it is also much more difficult to analyze. Further, because the neighbors in the Bayesian equilibrium of Jackson and Yariv (2007) are selected at random, the neighborhood structure looks like a random graph. Like other random-graph based models, there are typically multiple equilibria. On the other hand, in this paper, we are serious about the topology of the network and explain an important role of overlapping neighborhoods that cannot be captured in random graphs models.

The results of this paper are closely related to the literature on evolutionary learning and contagion in networks. Evolutionary game theory (Kandori *et al.* (1993), Young (1993), Blume (1993), Newton (2021), and many others) studies the long-run behavior of perturbed best response processes, where agents commit mistakes with a small probability, and instead of choosing a best response, take some other action.

A major contribution of this literature is a demonstration of a contagion phenomenon. Ellison (1993) (see also Ellison (2000)) shows that a best response may spread a risk-dominant action from a small initial set of deviators to the rest of a 1-dimensional lattice network. Blume (1995b) and Lee and Valentinyi (2000) extend this observation to higher-dimensional lattices. Morris (2000) describes general properties of networks for which Ellison's contagion wave exists. Morris (2000) also shows that risk-dominated actions cannot spread through a best response process regardless of the geometry of the network.

A strand of the literature studies evolutionary equilibrium selection in games with heterogeneous populations. For instance, Friedman (1991) describes a general framework with multiple continuum populations choosing actions and receiving payoffs and studies evolutionary steady states of continuous time adjustment dynamics. More closely related to this paper is Neary (2012), which studies a similar model to us but with two payoff shocks (more precisely, two subpopulations of deterministic size) and agents located on a complete graph. The paper presents conditions under which the evolutionary dynamics of Kandori *et al.* (1993) selects a fuzzy convention, i.e., an equilibrium where members of different subpopulations play different actions. Neary and Newton (2017) studies general payoff shocks and presents a sufficient condition under which the logit dynamics of Blume (1993) selects a fuzzy convention.

Our current results (specifically, Theorems 1 and 2) are related, but with some key differences. First, here, we are interested in static equilibria instead of a dynamic adjustment process. The evolutionary literature is subject to the criticism that one may need to wait for a very long time before reaching a stochastically stable outcome (Ellison (1993)). That criticism does not apply to our static model. Second, the previous papers study games with homogeneous payoffs and a behavior that is subject to small and disappearing perturbations: small and disappearing shocks in case of Ellison (1993) or Blume (1995b), and finite and small fraction of society modifying their actions in Lee and Valentinyi (2000) or Morris (2000). Instead, the payoff shocks in our model are significant, and, as a result, we are serious about heterogeneity. The non-trivial payoff shocks make our model more difficult to analyze, but they also render it closer to reality. Third, the evolutionary literature results show convergence to Harsanyi and Selten (1988)'s risk-dominance. Here, due to payoff heterogeneity, we need a new solution concept in the form of the RU-dominance. We show that the RU-dominance reduces to the risk-dominance when payoffs are homogeneous. Finally, the network topology plays an important role in both evolutionary models and in the current paper. In evolutionary models, the network affects the time for the coordination on risk-dominant outcome. However, it does not affect the final outcome: one of the key results of this literature is that risk-dominant coordination is (uniquely) stochastically stable on all networks (Peski (2010)). In our case, similarly to Lee and Valentinyi (2000) and Morris (2000), the network topology affects the equilibrium outcome.

In a recent contribution, Leister *et al.* (2022) study coordination games with fixed network and fixed (not random) threshold distribution. The paper works with arbitrary networks. To deal with a possible multiplicity of equilibria, they use global games as an equilibrium selection device. The authors develop an algorithm to compute the equilibrium adoption. The outcome of the algorithm depends on the details of payoff heterogeneity and how they interact with the topology of network. In contrast, in our paper, the assumption that thresholds are randomly and independently drawn from the same distributions allows us to separate the effects of the payoff distributions and the topology of the network.

2. Numerical example

Although our results are asymptotic, the coordination on RU-dominant outcome as well as the role of the networks can be demonstrated through simulations in a numerical example.

We compare the behavior under two threshold distributions. In both cases, the high action is strictly dominant for 30% of the population and the low action is strictly dominant for another 30%. Under P_1, the remaining 40% plays the high action only if at least 0.55 of their neighbors do the same. Under P_2, the remaining 40% plays the high action only if at least 0.4 of their neighbors do the same. The distributions are drawn in the top row of Figure 2.

If, like in Granovetter (1978), the population is continuum, and all agents play against the entire population, the equilibrium average behavior can be found as a fixed point of $P(.)$, i.e., an intersection of $P(.)$ with the 45°-line. In both cases, there are two stable equilibria: A with 0.3 and B with 0.7 fractions playing high. (In each case, there is also an unstable equilibrium in-between.) For each distribution, only one of these outcomes is RU-dominant - A in case of distribution P_1 and B in case of P_2.

Instead, consider a population of agents loving on one of two networks. Both networks have $\sim 60,000$ agents and each agent has, on average, ~ 120 neighbors.

- In a random graph (Erdős and Rényi (1959)), neighbors are randomly selected from the population.
- In a "city" network, people located on a two-dimensional grid. Each agent neighborhood is a square of agents of side equal to 11, centered at the agent.

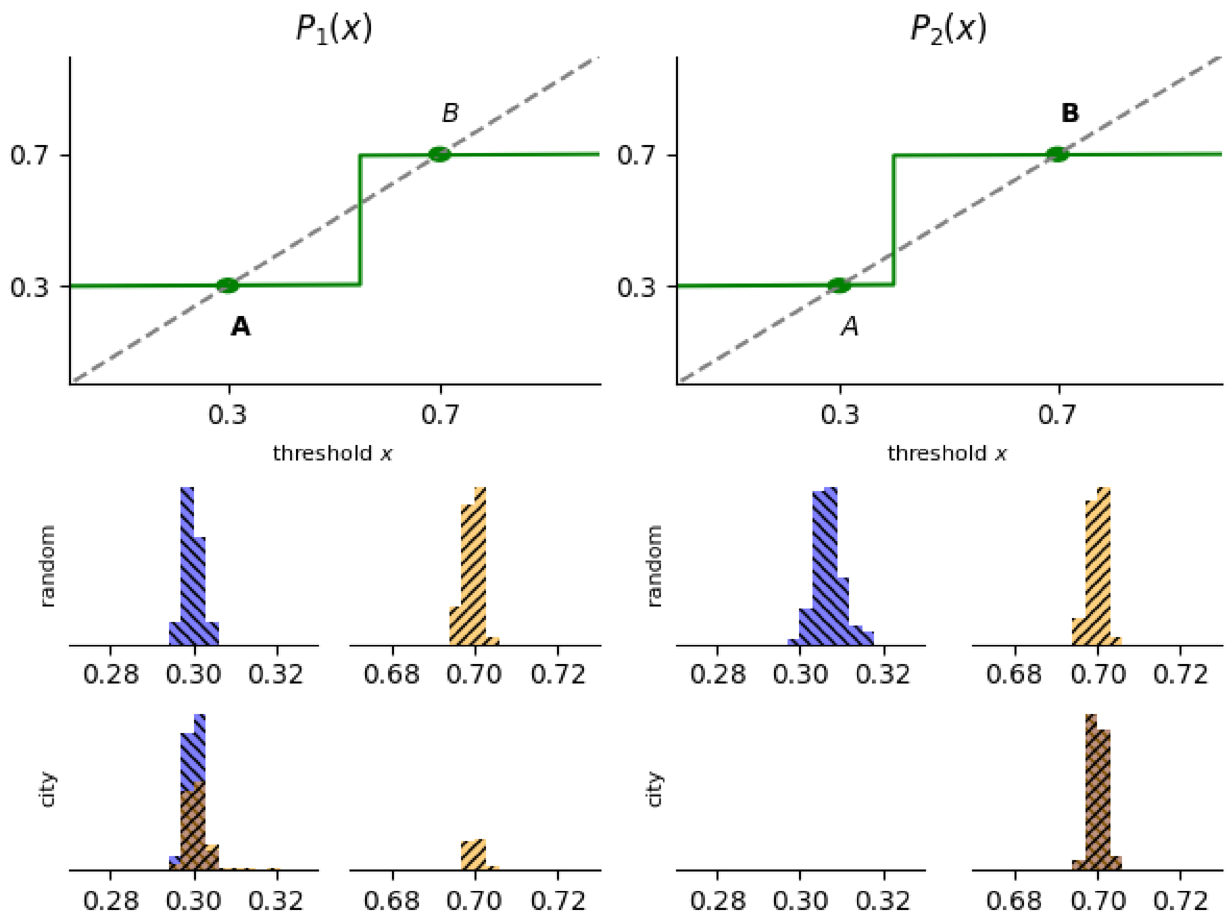


FIGURE 2. Monte-Carlo simulations of average equilibrium behavior in the lowest (blue, "\" hatch areas) and the highest equilibria (yellow, "/" hatch areas). The distributions substantially overlap (brown color) in the last row, corresponding to the city network.

We use Monte-Carlo simulations to estimate the probability distributions of average equilibrium behavior. In each simulation, we draw i.i.d. thresholds for all agents. For each realization of thresholds, we find the highest and lowest equilibria. Such equilibria are well-defined for binary coordination games. For example, to find the highest equilibrium, we start with a profile where all agents play the high action, and then run the best response process until none of the agents wants to change their action. Next, for each equilibrium, we compute the average equilibrium behavior. By combining average behaviors in two equilibria across different threshold realizations, we obtain the Monte-Carlo estimates.

These distributions for each network, each threshold distribution, and each equilibrium type (the lowest is marked with blue, "\" hatch areas and the highest with yellow, "/" hatch areas) are plotted in the two bottom rows of Figure 2. Because both distributions are highly concentrated around 0.3 (i.e., A) and 0.7 (i.e, B) values, for clarity, we only show regions around these two values.

There is a significant difference between random and city networks. In the random graph, the lowest and the highest equilibria correspond to the lowest (A) and highest (B) equilibria from the population model of Granovetter (1978), regardless of the threshold distribution. This is not unexpected as random graph with a relatively large number of agents is a good approximation of the continuum model.

On the city network, the range of equilibrium behaviors is much smaller and it depends on a threshold distribution. Under P_1, the lowest and the majority of realizations of the highest equilibria are concentrated around A. Under P_2, the average behavior in the highest and lowest equilibria is essentially equal to B. In other words, for a significant majority of threshold realizations, all equilibria on the city network has aggregate behavior consistent with the RU-dominant prediction.

The goal of the rest of the paper is to explain this pattern.

3. Model

3.1. **Model.** We study agents living in the nodes of a network. The network is defined as an undirected weighted graph with weights $g_{ij} = g_{ji} \geq 0$ for $i, j \leq N_g$, where N_g is the size of the network. The weights can be interpreted as a frequency of interactions between two agents and we assume that $g_{ii} = 0$. Let $g_i = \sum_j g_{ij} > 0$ for each agent i. Each agent i has a threshold τ_i drawn i.i.d. from probability distribution P. Each network g, and each realization of thresholds τ defines a complete information static game $G(g, \tau)$.

Each agent chooses a binary action $a_i \in \{0, 1\}$ and uses it in each interaction. The payoff in interaction with agent j is equal to $u_i(a_i, a_j, \tau_i) = a_i a_j - a_i \tau_i$, and the total payoff of agent i in all (weighted) interactions is equal to $\sum_j g_{ij} u_i(a_i, a_j, \tau_i)$. For each action profile a, let $\beta^a = (\beta_i^a)$ be a profile of average neighborhood fractions of agents who play action 1, i.e., $\beta_i^a = \frac{1}{g_i}\sum_j g_{ij}a_j$. An action profile is a Nash equilibrium if all agents best respond, or alternatively, if each agent plays action 1 (resp. 0) if the average action in their neighborhood is strictly larger (resp., smaller) than their threshold, i.e., for each i,

$$1\left\{\tau_i < \beta_i^a\right\} \leq a_i \leq 1\left\{\tau_i \leq \beta_i^a\right\}. \tag{2}$$

The model is strategically equivalent to general random-utility binary-action coordination games on networks.[1] The notion of equilibrium is a standard, static equilibrium of a complete information game. Although it is convenient to assume that agents know the thresholds and the network structure of the entire society, this assumption is neither realistic nor necessary. For the interpretation of the equilibrium, it is sufficient that agents observe the actions of their neighbors. Because ours is a coordination game, we can safely think about an equilibrium as a steady state of myopic best response adjustment process.

Two special cases are worth mentioning:

- homogeneous payoffs: Suppose that $\tau_i = \tau$ for all agents i (i.e., P is degenerate - see Figure 3 for $\tau = 0.4$ and 0.6). This is a standard model of coordination game on networks (Ellison (1993), Blume (1993), Lee and Valentinyi (2000), and others). Peski (2010) showed that, regardless of the network, various evolutionary dynamics select coordination on risk-dominant action as stochastically stable outcome,
- complete graph: Suppose that $g_{ij} = 1$ for each $i \neq j$. In the continuum limit $N_g \to \infty$, our model becomes equivalent to Granovetter (1978). Complete graph share similar features with Erdos-Renyi style random graphs.

We assume that none of the agents has significantly more connections than others, $\max_{i,j} g_i/g_j \leq w^*$, where $w^* < \infty$ is an (arbitrary) constant fixed throughout the

[1]A general model is as follows: For each agent i and j, i's payoff from interaction with agent j is equal to $u(a_i, a_j, \varepsilon_i)$, where $a_i, a_j \in \{0,1\}$ are actions and ε_i is a random shock to agent i's utility drawn from some distribution F. Assume that, for each ε,

$$\Delta(\varepsilon) := u(1,1,\varepsilon) + u(0,0,\varepsilon) - u(1,0,\varepsilon) - u(0,1,\varepsilon) > 0.$$

In order to translate this model to the threshold model, for each x, let $\tau_i = \frac{1}{\Delta(\varepsilon)}\left(u(1,0,\varepsilon_i) - u(0,0,\varepsilon_i)\right)$.

paper. This paper is concerned with asymptotic results when the network becomes sufficiently large and fine.

3.2. **Fuzzy convention.** For $\varepsilon > 0$ and $x \in [0,1]$, a profile a is *ε-fuzzy convention x* if all but ε fraction of agents play 1 if and only if their threshold is strictly below x:

$$\frac{1}{N_g} \left| a_i - \mathbf{1}\left(\tau_i \leq x\right) \right| \leq \varepsilon. \tag{3}$$

In a fuzzy convention, the behavior of almost all agents can be deduced from their threshold alone. Denote the 0-fuzzy convention of x as a profile a^x, where, for each agent i, $a^x = \mathbf{1}\left(\tau_i \leq x\right)$.

A fuzzy convention allows for a substantial heterogeneity of the behavior on micro level: individuals with different thresholds may choose different actions. At the same time, almost all agents use approximately the same procedure of determining their action as a function of their thresholds. Because the thresholds are i.i.d., for any large group of agents, with a large probability, fraction $P\left(x\right)$ of them will play 1. Thus, fuzzy conventions do not exhibit macro-level heterogeneity, with differences of aggregate behavior across different parts of the network.

An important feature of fuzzy convention is that it is network independent: (almost) all agents choose their actions purely based on their own threshold. Such strategies are necessarily employed in models like Jackson and Yariv (2007) and Galeotti *et al.* (2010), where agents choose their action *before* their neighbors are drawn randomly from the rest of the population. In principle, there is no reason why such strategies should play any role in our paper, which is concerned with static complete information equilibrium and where actions must be best responses *after* the network and all thresholds are determined.

In a fuzzy convention x, the expected fraction of neighbors who play 1 is equal to $P\left(x\right)$. It turns out that if the agents' neighborhoods are sufficiently large, the expected fraction is also close to the observed one. If the fuzzy convention is also an equilibrium, we expect that that $x \sim P\left(x\right)$.

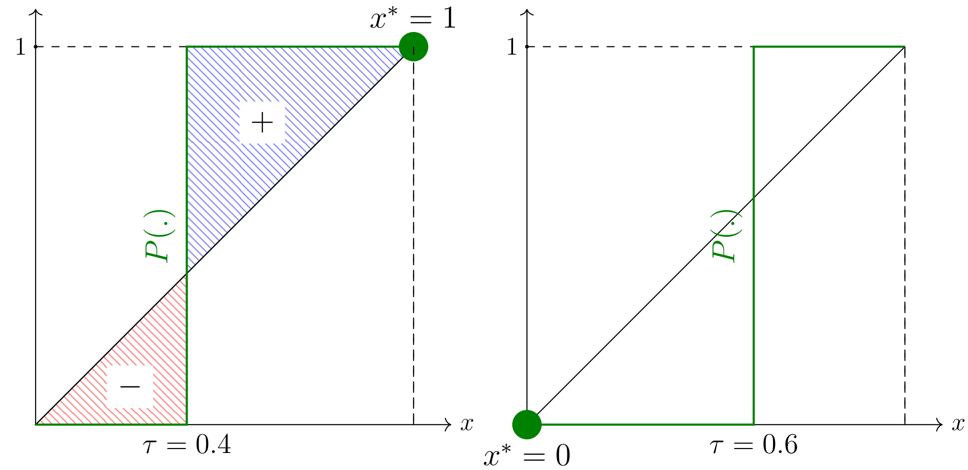


FIGURE 3. RU- and risk-dominance when P is degenerate for $\tau = 0.4$ (left panel) and $\tau = 0.6$ (right panel).

3.3. *RU*-**dominant outcome.** An outcome $x^* \in [0,1]$ is *random utility (RU) dominant* if

$$x^* \in \arg\max_x \int_0^x \left(y - P^{-1}(y)\right) dy. \tag{4}$$

(When P is not invertible, we define $P^{-1}(y) = \inf\{(x : P(x) \geq y)\}$.) It is *strictly RU-dominant*, if it is the unique maximizer. Graphically, the integral (4) is equal to the sum of signed measures of areas between the cdf $P(.)$ and the 45° line: the area below the 45° line and above $P(.)$ is added with a "$-$" sign and the area above the 45° line and below $P(.)$ is added with the "$+$" sign. Fig. 1 illustrates such a calculation generic function $P(.)$.

Any maximizer of (4) is a non-atomic fixed point of $P(x) = x$. However, even if there are multiple stable fixed points, generically, there exists a unique RU-dominant outcome.

In a special case of homogeneous payoffs (see Section 3.1), the definition of RU-dominance reduces to the risk-dominance of (Harsanyi and Selten (1988)). To see that, suppose that $P(.)$ is degenerate and concentrated on a single threshold τ (i.e., there is no uncertainty about thresholds). Figure 3 shows the distribution P for two values of τ. In both cases, the integral from expression (4) is equal to $\frac{1}{2}x^2 - \tau x = x\left(\frac{1}{2} - \tau\right)$ and

- when $\tau = 0.4$, the integral is maximized at $x^* = 1$,

- when $\tau = 0.6$, the integral is maximized at $x^* = 0$.

In both cases, the RU-dominant outcome is identical to the risk-dominant one.

For future reference, note that any strictly RU-dominant outcome is also a unique maximizer of

$$\nu(x) = \frac{1}{2}(P(x))^2 - \int_0^x y dP(y). \tag{5}$$

Indeed, the maximizer of (5) must satisfy $P(x) = x$, and a change of variables shows that the two expressions are equal for such x.

4. RU-dominant fuzzy convention

This section contains the first main result of the paper: all sufficiently large and fine networks have an equilibrium that is fuzzy convention x^*. Define a bound on the importance of a single agent in another agent's neighborhood as

$$d(g) = \max_{i,j} \frac{g_{ij}}{g_i} \in [0, 1].$$

For $d(g)$ to be small, each agent must have many neighbors. In the next result, the phrase "with probability" refers to the probability distribution over all threshold profiles:

Theorem 1. *Suppose that x^* is the strictly RU-dominant outcome. For each $\eta > 0$, there is $d > 0$ such that, for each network g st. $d(g) \le d$, with probability $1 - \eta$, there is an equilibrium that is η-fuzzy convention x^*.*

If the network is sufficiently fine, i.e. when $d(g)$ is small, then, for almost all realizations of thresholds, there is an equilibrium where almost all agents behave as if they were playing fuzzy convention x^*.

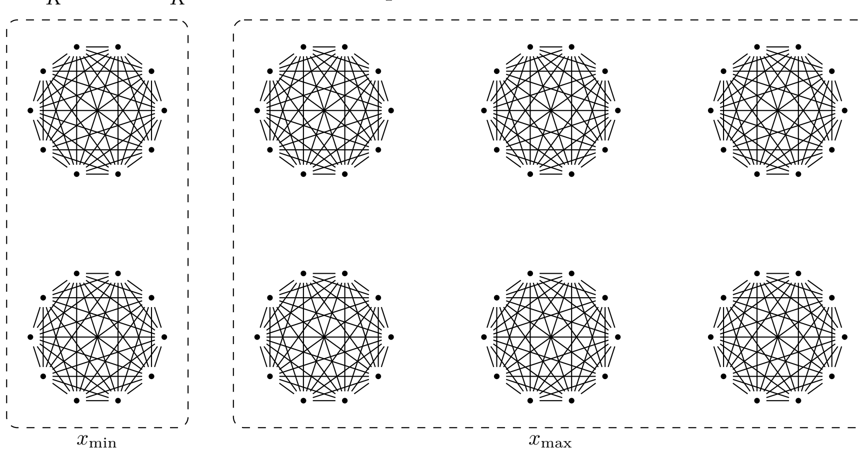
It is worth pointing out that the Theorem is not true for any other $x \neq x^*$. This is because there are networks on which there are no other equilibria than fuzzy conventions x^* (see Theorem 2 below). The fact that the result does not hold for any other fixed point of $P(.)$ but x^* should caution the reader that there is no "straightforward" or "immediate" proof based on Granovetter (1978) or convergence to random graphs.

At the same time, the Theorem does not say that that the equilibrium is unique, or that all equilibria are fuzzy conventions x^*, or even that all equilibria are fuzzy conventions. None of it is true. For example, one can easily show that the Granovetter's analysis of the continuum population extends to complete or random graphs: If the population is sufficiently large, for a large probability set of thresholds realizations, complete graph has multiple equilibria, including x-fuzzy convention for each x that is a fixed point of $P(.)$. Even more, denoting by $x_{\min}$ and $x_{\max}$ the smallest and the largest fixed points of $P(.)$, there are networks, where, with a large probability, there is an equilibrium with average behavior close to x for each $x \in [x_{\min}, x_{\max}]$ and there are many equilibria that are not fuzzy conventions.[2]

The condition that $d(g)$ is small corresponds to requirements of vanishing influence in social learning literature (for instance, Jackson (2010) or Mossel *et al.* (2015)). In our case, we require that each agent has a vanishing influence on every other agent. This ensures that the empirical (i.e., realized) distribution of thresholds in each agent neighborhood weighted by the link weights is close to distribution $P(.)$. This is related to the role of this condition in social learning literature, where it is important that a random signal observed by one agent does not unduly affect the rest of the society.

The proof of Theorem 1 relies on the fact that the threshold model is a potential game (Monderer and Shapley (1996)). For each action profile a and threshold profile

[2]An example is a network consisting of K complete graphs, each of size N, disconnected from each other. Then, for each $k \leq K$, if N is sufficiently large, there is an equilibrium where agents in the first k complete graphs play fuzzy convention $x_{\min}$ (restricted to this complete graph) and agents in the remaining $K-k$ complete graphs play fuzzy convention $x_{\max}$. Such a profile is an equilibrium, but it is not a fuzzy (or any other) convention in the network as a whole. The average behavior in such a profile is $\frac{k}{K}x_{\min} + \frac{K-k}{K}x_{\max}$. An example where $K = 8$ and $k = 2$ can be found below.



τ, define

$$V(a;\tau) = \frac{1}{2}\sum_{i,j} g_{ij}a_ia_j - \sum g_ia_i\tau_i. \tag{6}$$

Then, $V(a_i, a_{-i};\tau) - V(a_i', a_{-i};\tau) = g_i(\beta_i^a - \tau_i)(a_i - a_i')$, which implies that $V(1, a_{-i};\tau) - V(0, a_{-i};\tau) \geq 0$ if and only if $\beta_i^a \geq \tau_i$, or if and only if 1 is a best response for agent i. In other words, V is an (ordinal) potential function. Monderer and Shapley (1996) shows that a profile is an equilibrium profile of a potential game if and only if it is a *local* maximizer of a potential function.

We emphasize that (6), not the expression in (4), is the potential of the game for a given network. The latter can be shown to be a (some type of) potential of the continuum limit of complete graphs. Formula (6) applies to all networks.

The proof consists of four steps. First, we show that, if the network is sufficiently large and fine, then, with a large probability, the potential of profile a^{x^*} is very close to the maximum value of (5). In the next two steps, we consider all profiles a such that a is an equilibrium that is not ε-fuzzy convention. In the second step, we show that, if inequality (3) fails, neighborhood averages β_i^a must be significantly different from x^*. Third, we estimate potential for all such profiles a and show that it is approximately equal to the value that depends on β_i^a and such that is strictly smaller than the maximum of (5) for $\beta_i^a \neq x^*$. Finally, we recall that any maximizer of the potential must be an equilibrium. Together with the previous steps, this observation imply that the maximizer must be a fuzzy convention x^*.

The fourth step is immediate. The first step is a relatively straightforward application of a standard concentration inequality (i.e., a version of the law of large numbers). The second and the third step are relatively straightforward calculations that rely on a version of the concentration inequality that holds uniformly across all profiles a. The proof of the latter is the most difficult part of the entire argument. The reason why we need an uniform concentration inequality is that the bounds used in computations in the second and the third step must simultaneously hold for all profiles a. [3]

[3]To see the difference between the two types of probabilistic inequalities, consider the following problem. Suppose that $a_i, \tau_i \in \{0,1\}$, and τ_is are i.i.d., uniformly distributed on $\{0,1\}$. Consider a function $V_0(a,\tau) = \frac{1}{N}\sum a_i\tau_i$. Then, for each arbitrarily small $\varepsilon > 0$, there exists N_ε such that for each $N > N_\varepsilon$,

$$\sup_a P\left(\left|V_0(a,\tau) - \frac{1}{2}\right| > \varepsilon\right) < \varepsilon.$$

4.1. **Concentration inequalities.** We sketch the main steps of the proof. We start with a concentration inequality. Let $\mathcal{F}$ be the set of measurable functions $f:[0,1]^2 \to [0,1]$. For each $f \in \mathcal{F}$, each b, let $\mathbb{E} f(.,b) = \int f(x,b)\,dP(x)$ denote the expectation of $f(.,b)$ with respect to the distribution of thresholds P. The Hoeffding inequality implies that there exists constants $B < \infty$ and $c_\varepsilon > 0$ such that for each profile a and measurable function $f(\tau,\beta) \in [0,1]$,

$$\text{Prob}\left(\left|\sum_i g_i f(\tau_i,\beta_i^a) - \sum_i g_i\,\mathbb{E} f(.,\beta_i^a)\right| \geq \varepsilon \sum g_i\right) \leq B\exp\left(-c_\varepsilon N_g\right). \tag{7}$$

(Here, and below, Prob is the probability over the realizations of threshold profiles.) Similarly, the Hanson-Wright inequality says that, for possibly different constants B and c_ε,

$$\text{Prob}\left(\left|\sum_{i,j} g_{ij}\left(\prod_{k=i,j} f(\tau_k,\beta_k^a)\right) - \sum_{i,j} g_{ij}\left(\prod_{k=i,j}\mathbb{E} f(.,\beta_k^a)\right)\right| \geq \varepsilon \sum g_i\right) \leq B\exp\left(-c_\varepsilon N_g\right). \tag{8}$$

The above inequalities hold for each profile a separately. The next Lemma show that they can be strengthened to hold *uniformly* across all profiles.

Lemma 1. *There exist constants $B < \infty$ and $c(\varepsilon,K,d)$ for each $\varepsilon > 0$, $K < \infty$, and $d > 0$ such that $\liminf_{d\to 0} c_{\varepsilon,K,d} > 0$ and such that if $f \in \mathcal{F}$ is a K-Lipschitz function, then*

$$\begin{aligned}&\mathit{Prob}\left(\sup_a\left|\sum_i g_i f(\tau_i,\beta_i^a) - \sum_i g_i\,\mathbb{E} f(.,\beta)\right| \geq \varepsilon \sum g_i\right)\\ \leq& B\,\mathit{exp}\left(-c_{\varepsilon,K,d(g)} N_g\right),\end{aligned} \tag{9}$$

$$\begin{aligned}&\mathit{Prob}\left(\sup_a\left|\sum_{i,j} g_{ij}\left(\prod_{k=i,j} f(\tau_k,\beta_k^a)\right) - \sum_{i,j} g_{ij}\left(\prod_{k=i,j}\mathbb{E} f(.,\beta_k^a)\right)\right| \geq \varepsilon \sum g_i\right)\\ \leq& B\,\mathit{exp}\left(-c_{\varepsilon,K,d(g)} N_g\right).\end{aligned} \tag{10}$$

At the same time, the uniform version of the above inequality is not valid: for any N,

$$P\left(\sup_a\left|V_0(a,\tau) - \frac{1}{2}\right| > \varepsilon\right) > 1-\varepsilon.$$

In fact,$\sup_a\left|V(a,\tau) - \frac{1}{2}\right| = \frac{1}{2}$ and it can be attained by $a = \tau$.

The proof of the Lemma establishes probabilistic bound (9) (resp., (10)) as a product between bound (7) (resp., (8)) and a measure of the size of the set of neighborhood profiles $\mathcal{B} = \{\beta^a : a \text{ is a profile}\}$. To explain the idea, notice that, for any function $F(\beta^a)$ of the neighborhood profile β^a, we get:

$$\begin{aligned}\text{Prob}\left(\sup_a F(\beta^a)\right) &= \text{Prob}\left(\sup_{\beta\in\mathcal{B}} F(\beta)\right)\\ &\leq |\mathcal{B}| \sup_{\beta\in\mathcal{B}} \text{Prob}\left(F(\beta)\right) = |\mathcal{B}| \sup_a \text{Prob}\left(F(\beta^a)\right).\end{aligned}$$

In other words, the uniform probabilistic bound is a product of the individual bound and the counting measure of set $\mathcal{B}$. It turns out that the counting measure is too large ($|\mathcal{B}| \sim \exp(2N_g)$) for our purposes. Instead, the proof relies on the metric entropy of set $\mathcal{B}$ (see Appendix A.1 for details). We show the metric entropy of $\mathcal{B}$ is of order $\exp(d(g)N)$, which, when $d(g)$ is small, leads to bounds that are sufficient to conclude the proof of Lemma 1. The use of metric entropy requires some modifications to the above argument, including the restriction to Lipschitz functions f.

4.2. **Estimates of the potential function.** We use Lemma 1 in three calculations below. In all cases, we assume that the network is sufficiently large and fine and the thesis of the Lemma holds. First, we find the potential of 0-convention x^* profile $a^* = a^{x^*}$: for each i, $a_i^* = \mathbf{1}\{\tau_i \leq x^*\}$.

Note that $\mathbb{E}\,\mathbf{1}\{. \leq x^*\} = P(x^*)$. Lemma 1 implies the following estimate:

$$\begin{aligned}V(a^*;\tau) &= \frac{1}{2}\sum_{i,j} g_{ij}a_ia_j - \sum g_ia_i\tau_i\\ &= \frac{1}{2}\sum_{i,j} g_{ij}\mathbf{1}\{\tau_i \leq x^*\}\mathbf{1}\{\tau_j \leq x^*\} - \sum g_i\mathbf{1}\{\tau_i \leq x^*\}\tau_i\\ &\approx \frac{1}{2}\sum_{i,j} g_i(P(x^*))^2 - \sum g_i\int_0^{x^*} ydP(y) = \sum g_i\nu(x^*).\end{aligned}$$

(Because $\mathbf{1}\{. \leq x^*\}$ is not Lipschitz, the Lemma is applied to a Lipschitz approximation - the details are left for the Appendix).

Second, take an arbitrary equilibrium profile that is not ε-fuzzy convention of x^*. Because of (2) and (3), we get

$$\begin{aligned}\varepsilon &\leq \frac{1}{N_g}\left|a_i - \mathbf{1}\left(\tau_i \leq x^*\right)\right| \\ &\leq \frac{1}{N_g}\sum_i \Big(\mathbf{1}_{\beta_i^a \leq x^*}\mathbf{1}\left(\tau_i \in [\beta_i^a, x^*]\right) + \mathbf{1}_{\beta_i^a \geq x^*}\mathbf{1}\left(\tau_i \in [\beta_i^a, x^*]\right)\Big).\end{aligned}$$

By Lemma 1, with a large probability, the following bound holds:

$$\frac{1}{N_g}\sum_i \left|P\left(\beta_i^a\right) - P\left(x^*\right)\right| \geq \frac{1}{2}\varepsilon. \tag{11}$$

Third, we estimate the potential for such a profile a. Applying Lemma 1 once more, we obtain

$$\begin{aligned}V\left(a;\tau\right) &= \frac{1}{2}\sum_{i,j} g_{ij}a_ia_j - \sum g_ia_i\tau_i \\ &= \frac{1}{2}\sum_{i,j} g_i\mathbf{1}\left(\tau_i \leq \beta_i^a\right)\mathbf{1}\left\{\tau_j \leq \beta_j^a\right\} - \sum g_i\mathbf{1}\left(\tau_i \leq \beta_i^a\right)\tau_i \\ &\approx \frac{1}{2}\sum_{i,j} g_{ij}P\left(\beta_i^a\right)P\left(\beta_j^a\right) - \sum g_i\int_0^{\beta_i^a} ydP\left(y\right).\end{aligned}$$

Because $2P\left(\beta_i^a\right)P\left(\beta_j^a\right) \leq P\left(\beta_i^a\right)^2 + P\left(\beta_j^a\right)^2$, the potential of a is not larger than

$$\leq \frac{1}{2}\sum_{i,j} g_{ij}\left(P\left(\beta_i^a\right)\right)^2 - \sum g_i\int_0^{\beta_i^a} ydP\left(y\right) = \sum_i g_i\nu\left(\beta_i^a\right).$$

By the remark at the end of Section 3.3, unless $\beta_i^a = x^*$, the above is strictly smaller than the potential of a^*. Hence, together with the estimate of potential for profile a^*, the bound (11) implies that an arbitrary equilibrium profile that is not ε-fuzzy convention of x^* cannot maximize potential.

Finally, recall that any potential maximizer must be an equilibrium. It follows that the potential maximizer must be ε-fuzzy convention of x^*.

5. RU-dominant selection

In the previous section, we showed that all sufficiently fine networks have equilibria that are fuzzy conventions x^*. Here, we show that there are networks where, with a large probability, all equilibria are fuzzy conventions x^*:

For each $\eta > 0$, the proof constructs a "city" network, where agents live on a 2-dimensional grid and interact with other agents who live around them. The network is parameterized with M and m. There are M^2 agents living on square $\left[0, \frac{M}{m}\right]^2 \subseteq \mathbb{R}^2$ at fractional points $\left(\frac{k}{m}, \frac{l}{m}\right)$ for $k, l = 1, ..., M$. Any two agents i and j are connected, $g_{ij} = 1$, if the (Euclidean) distance between them is no larger than 1. To avoid separately dealing with border cases, we assume that all distance calculations are done $\mathrm{mod}\frac{M}{m}$, which transforms the square $\left[0, \frac{M}{m}\right]^2$ into a torus.

Theorem 2. *Suppose that x^* is the strictly RU-dominant outcome and that either (a) $x^* \in (0,1)$ and $0 < P(0) \leq P(1) < 1$, (b) $x^* = 1$ and $P(0) > 0$, or (c) $x^* = 0$ and $P(1) < 1$. For each $\eta > 0$, if m and $\frac{M}{m}$ are sufficiently large, then with probability $1 - \eta$, each equilibrium on (M, m) city network is η-fuzzy convention x^*.*

The Theorem says that there exist networks, where all equilibria are fuzzy conventions x^*, or that all equilibria have a form identified by Theorem 1.

We emphasize that the Theorem makes a statement about *static*, *complete information* game equilibria. At the same time, the proof relies on a *dynamic* technique of contagion waves (Ellison (1993); Morris (2000)). We show that if an action profile is, in some sense, higher (resp., lower) than fuzzy convention x^*, then best response dynamics will push the profile below (resp., above) x^*. This shows that the original profile could not have been an equilibrium. We describe the intuition behind the proof, including the relation to the maximization problem, below.

If $P(0) > 0$ (resp., $P(1) < 1$), then, with a positive probability, there are agents for whom action 1 (resp., 0) is strictly dominant and it is played in any equilibrium. The only assumption of the Theorem is that there is a positive probability of such agents. The role of such agents is similar to the role of initial infectors in Lee and Valentinyi (2000) and Morris (2000) or the role of small probability mistakes in evolutionary models.

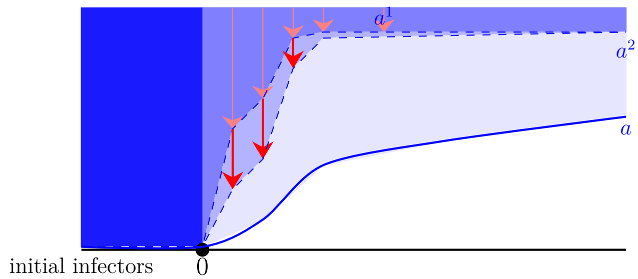


Figure 4. Contagion wave

The city network is an example of a 2-dimensional lattice. The proof could easily extend to $K > 2$ dimensional lattices (but, as we explain below, not to $K = 1$). After we describe the proof, we point to the properties of multi-dimensional lattices that are important for the proof. Extending the Theorem to other networks is beyond the goals of this paper.

5.1. **Contagion on line.** Next, we describe the intuition for the proof. We assume that $x^* = 0$ and $P(1) < 1$.

We start with the intuition behind the contagion argument. It is useful initially to work with a toy version of the line network from Ellison (1993) (the general argument does not work on line and it requires at least 2-dimensional lattices). Suppose that agents are distributed uniformly along a line at discrete and equally spaced locations. Each location contains a continuum population of mass 1. The populations in locations i and j are connected with each other, with weights that depend only on the distance $g_{ij} = g_{i-j} =: g_{j-i}$. We assume there are no connections between agents in the same location, i.e., $g_0 = 0$, and the weights are normalized so that $\sum g_d = 1$. Finally, we assume that there are no connections between agents at distance larger than d: $g_{i-j} = 0$ for $|i - j| > d$.

Take an action profile a^0 such that agents in locations $i \in [-2d, 0]$ play action 0 and all other agents play 1. In our model (but not its continuum toy version), assumption $P(1) < 1$ implies that there is a positive probability that a contiguous group of agents have 0 as a strictly dominant action. If the line network is long enough, the existence of a group of $2d$ agents who play 0 for sure can be guaranteed with a probability arbitrarily close to 1.

Going back to the toy line with continuum of agents in each location, consider a revision process in which agents in all locations apart from $i \geq 0$ switch to their myopic best responses. Complementarities imply that they can switch at most once, and if they do, they switch from action 1 to 0. Figure 4 illustrates the first two stages of such a process. In the first stage, actions are changed by agents in locations $i > 0$ for whom action 0 is strictly dominant action, as well as high-threshold agents in locations $i \in [0, d]$ for whom 0 is a best response given a^0. In the second stage, additional agents in locations $i \leq 2d$ may change actions. And so on. The process will continue until a stable point where no more agents $i \geq 0$ want to switch to 0. Denote the fraction of agents who play 1 in location i in stage n as a_i^n and the limit fraction as $\lim_n a_i^n = a_i$. Due to the payoff complementarities, profiles a_i^n for each n and a_i must be increasing in i.

In this toy version, the continuum law of large numbers allows us to express the fraction of agents for whom 1 is a best response given profile a as $P\left(\sum_d g_d a_{i+d}\right)$. Given that a is the limit of the best response dynamics, we have, for each location $i \geq -2d$,

$$a_i \leq P\left(\sum_d g_d a_{i+d}\right).$$

Taking the inverse, we obtain

$$P^{-1}(a_i) \leq \sum_d g_d a_{i+d} = \sum_j \left(\sum_{d \geq j-i} g_d\right)(a_{j+1} - a_j),$$

where the equality is due to a discrete version of the integration-by-parts formula and the fact that $a_i \geq 0$ for each i. After multiplying by $a_{i+1} - a_i \geq 0$, and summing up across all locations i, we get

$$\sum_i P^{-1}(a_i)(a_{i+1} - a_i) \leq \sum_{i,j} \left(\sum_{d \geq j-i} g_d\right)(a_{i+1} - a_i)(a_{j+1} - a_j). \tag{12}$$

The left-hand side of the inequality is approximately equal to $\int_0^a P^{-1}(y)\,dy$ when the distance between locations is small (which corresponds is guaranteed by large m assumption). To compute the right-hand side, notice that we can switch the roles of i and j in the summation without affecting its value. Together with the fact that

$\sum_{d\geq j-i} g_d + \sum_{d\geq i-j} g_d = \sum g_d = 1$, we get

$$\begin{aligned}
&\sum_{i,j}\left(\sum_{d\geq j-i} g_d\right)(a_{i+1}-a_i)(a_{j+1}-a_j)\\
=&\frac{1}{2}\left(\sum_{i,j}\left(\sum_{d\geq j-i} g_d+\sum_{d\geq i-j} g_d\right)(a_{i+1}-a_i)(a_{j+1}-a_j)\right)\\
=&\frac{1}{2}\left(\sum_{i,j}(a_{i+1}-a_i)(a_{j+1}-a_j)\right)=\frac{1}{2}a^2\\
=&\frac{1}{2}a^2=\int_0^a ydy.
\end{aligned}$$

Putting the two sides together, inequality (12) implies that

$$\int_0^a \left(y-P^{-1}(y)\right)dy \geq 0.$$

If $a>0$, this contradicts the fact that $x^*=0$ is the unique maximizer of the integral on the right-hand side of (4). Thus, in the limit of best response revision process, it must be that all locations play $a_i=0$.

The contagion argument extends from a line to higher-dimensional lattices due to an elegant argument from Blume (1995b) (see also Lee and Valentinyi (2000) and Morris (2000)). The idea is that if the initial group is sufficiently large, we can approximate it using a set with a smooth (i.e., low curvature) boundary. Then, we can analyze the spread of the contagion wave behavior in the direction that is normal to the boundary. This trick turns the problem into a one-dimensional one, and the above argument applies.

5.2. **Obstacles.** Although the continuum assumption is useful in explaining the intuition, the argument needs to be modified for our model. For example, the assumption ignores a positive probability of a contiguous group of “bad” agents for whom 1 is the strictly dominant action. If sufficiently large, such a group of “bad” agents will stop the best response revisions towards action 0 and block the contagion wave (see the left panel of Figure 5).

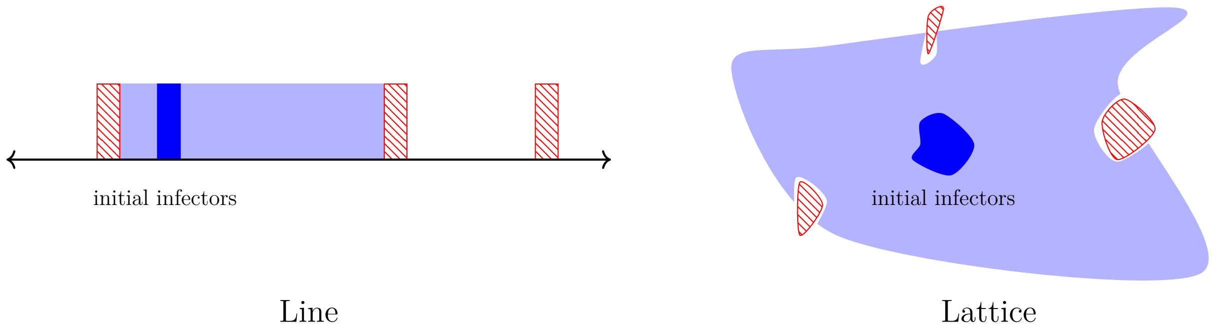


FIGURE 5. Obstacles to the contagion wave

"Bad" sets cannot be eliminated or avoided in the one-dimensional "line" network. However, "bad" sets are intuitively less likely to block the contagion wave on higher-dimensional lattices (see the right panel of Figure 5). The reason is that to block the wave, the "bad" sets would have to be arranged so as to surround it. We show that, on a two-dimensional lattcice, if m and $\frac{M}{m}$ are sufficiently large, the likelihood of "bad" sets surrounding the initial infectors is very small.

5.3. **Proof summary.** More generally, without the continuum assumption, the argument behind contagion waves must work with finite laws of large numbers. Below, we sketch the main ideas how we do it. The details of the proof can be found in Appendix B.

The lattice is divided into large and small cubes so that the number of large cubes in the lattice is very large, each large cube contains a very large number of disjoint neighborhoods, each neighborhood contains a very large number of small cubes, and each small cube contains a very large number of agents (see Figure 6). These numbers are chosen so that the following series of claims holds:

(1) The number of agents in a small cube and the number of small cubes in a neighborhood are sufficiently large, so that the fraction of shared agents and the fraction of shared small cubes in neighborhoods of any two agents i and j is well approximated by the area of the intersection of two 1-radius circles with centers at i and j (Lemma 3).
(2) The size of each small cube is sufficiently large so that, for each small cube, with a probability close to 1, the empirical distribution of payoff shocks within

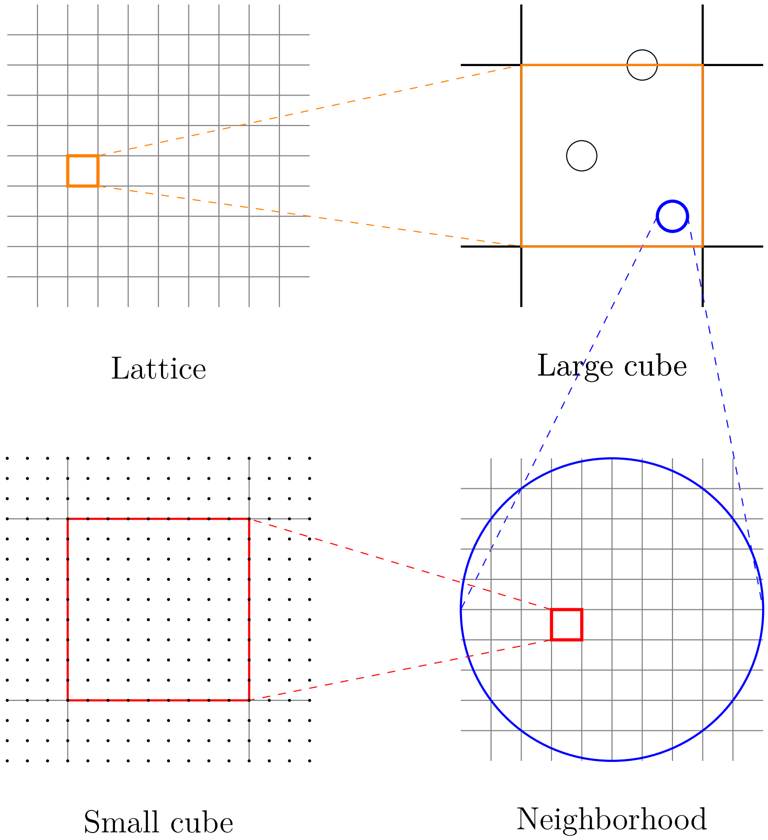


FIGURE 6.

the cube is close to the true distribution. We say that a small cube is *(γ-)bad* if, for some fraction x, the average best response action of the agents within the cube is (γ-)larger than $P(x)$. Agents in bad cubes may tilt toward higher best responses than a statistical agent. Agents in a small cube that is not bad are well approximated by the continuum assumption in the following sense: the average best response in the small cube is not higher than $P(\beta)$, where β is the average "belief" (i.e., the average neighborhood action) for members of the cube.

(3) A large cube is *good* if it contains no bad small cubes. The ratio of the size of a small cube (i.e., the number of agents within each small cube) to the number of small cubes in a large cube is sufficiently large, so that the probability p that the large cube is good is arbitrarily close to 1.

A large cube is *extraordinary*, if it contains only agents for whom 0 is the strictly dominant action. Extraordinary cubes play the role ofinitial infectors. The number of large cubes is sufficiently large, so that the probability that an extraordinary large cube exists is arbitrarily close to 1.

(4) Two large cubes are *connected* if they share a wall. The number of large cubes is sufficiently large, and the probability p that a large cube is not good is sufficiently small, so that there exists a giant component of good large cubes - a set of good large cubes that contains almost all large cubes on the lattice and such that all of its elements are connected with each other by paths of good large cubes that share a wall. This argument is the content of Lemma 7 and it relies on definitions and results from the percolation theory (Bollobás *et al.* (2006)).
   (a) First, we show that each connected set S can be surrounded by a connected "boundary" ∂S that isolates set S (and, possibly, some other large cubes) from the remaining large cubes. The total number of large cubes isolated away from set S is not larger than $|S|^2$. (On a two-dimensional lattice, the worst-case scenario bound comes from elements of set S arranged in a way that surrounds an interior proportional in size to the square of its perimeter.)
   (b) For a collection of connected sets $S_1, ..., S_J$ that are *not* connected with each other, the giant connected component that omits all sets S_j contains all but at most $\sum |S_j|^2$ large cubes.
   (c) Let $S_1, ..., S_J$ be the collection of all maximally connected collections of large *bad* cubes. We estimate the expected value of $\sum |S_j|^2$ as proportional to the number of all large cubes multiplied by the probability p that a single large cube is bad (Lemma 5). An application of the Markov inequality shows that, if p is sufficiently small, the giant connected component that contains only good cubes contains a fraction of all large cubes that is arbitrarily close to 1.

(5) Using the ideas from Blume (1995b), we show that if the curvature of the two-dimensional contagion wave is sufficiently small relative to the curvature of an

individual neighborhood, the contagion wave will spread, as long as its path contains only good small cubes (Lemma 9).

Putting it together, the contagion wave is going to spread through a vast majority of the giant connected component of good large cubes, and thus a vast majority of the lattice. Hence, with a large probability, the average action in the largest equilibrium on a sufficiently large two-dimensional lattice is close to x^*.

5.4. **Key properties of the city network.** We summarize the above discussion by identifying four properties of (M, m)-city network that play key role in the proof.

(1) *Large number of connections* m allows us to approximate the empirical distribution of thresholds in an agent's neighborhood by the model distribution P. This approximation forms a basis for the continuum model discussed in section 5.1.
(2) *Large network*: The population must be sufficiently large to ensure that, for each action, with a high probability, there is a sufficiently large number of agents for whom this action is strictly dominant. Such agents start the contagion argument and they play a similar role as initial infectors in Lee and Valentinyi (2000) or Morris (2000). In the city network, we require that $\frac{M}{m}$ is sufficiently large.
(3) *Slow neighborhood growth*: For the contagion argument of section 5.1 to hold, the size of neighborhoods must grow sufficiently slowly (see Morris (2000) for the definition and properties).
(4) *Percolation property*: The contagion cannot be obstructed by the obstacle phenomenon described in section 5.2. Using the language introduced above, the good set of cubes must contain a large connected component of the graph.

It is not immediately obvious how to formalize the last property in a simple way. (A non-simple way is to assume that the thesis of Lemma 4 from the Appendix must hold.) We leave this task for future research.

6. Equilibrium selection

In this section, we point out two equilibrium selection theories that select fuzzy convention x^* as the unique solution for random utility coordination games on networks.

6.1. **Evolutionary stability.** The proof of Theorem 1 shows that fuzzy convention x^* is, with a large probability, a global maximizer of a potential function for the coordination game. Recall that global maximizers of the potential function are selected in complete information static coordination games by two different equilibrium selection theories: robustness to incomplete information (Ui (2001)) and stochastic stability under logistic dynamics (Blume (1993), Blume (2018)).

6.2. **Robust behavior.** Next, we explain that fuzzy convention $x^*$is the only behavior that is robust to incomplete information about the network. The idea is parallel to the definition of robustness to incomplete information from Kajii and Morris (1997). We take a perspective of a researcher/analyst who observes a large population of agents and attempts to predict individual behavior $\alpha_i(\tau_i) \in [0,1]$, where $\alpha_i(\tau)$ is the probability of playing action 1, as a function of individual thresholds τ_i. The researcher understands that the agents play a coordination game with their neighbors on some large and fine network and understands the parameters of the game, but they do not necessarily understand the details of the network topology. She would like her prediction to be robust to a mis-specification of the network.

Definition 1. A threshold behavior $(\alpha_i(.))_i$ is *robust to the misspecification of the network* if and only if, for each η, there exists $d > 0$, such that for each network g, if $d(g) < d$, with probability at least $1-\eta$ (over the realization of thresholds τ_i) , there exists an equilibrium $a_i \in \{0,1\}$ of the network game $G(g,\tau)$ such that

$$\frac{1}{N_g} \sum_{i \leq N_g} |a_i - \alpha_i(\tau_i)| \leq \eta.$$

To interpret the definition, notice that threshold behavior $(\alpha_i(.))_i$ is network-independent: each agent's action depends on their own threshold and not to whom they are connected, and what their neighbors are doing. If the behavior is robust to misspecification, it is an equilibrium for a great majority of agents, whatever is the true network of interactions, and whether the agents know the network or not. In other words, the behavior is approximately an equilibrium regardless whether the network is misspecified by the agents or the researcher.

Recall that the 0-fuzzy convention of x^* is a network-independent profile where agents play 1 if and only if their threshold is smaller than x, $a^*(\tau_i) = \mathbf{1}(\tau_i \leq x^*)$.

Theorem 3. *Suppose that x^* is the strictly RU-dominant outcome. Then, a threshold behavior α is robust to mis-specification of the network if and only if it is 0-fuzzy convention of x^*.*

The above result shows that playing a^* is the only profile that is robust to misspecification of the network.

Proof. "If" direction follows from Theorem 1. The "only if" direction follows from Theorem 2. □

7. Conclusions

This paper presented a theory of behavior in random utility binary coordination games on large networks. We showed that on some networks, with a large probability, large coordination games have essentially a unique equilibrium. Because this equilibrium exhibits micro-, but not macro-level heterogeneity of behavior, we refer to it as a fuzzy convention. The average behavior in such a convention corresponds to a natural extension of risk-dominance from deterministic to random-utility coordination games. We also showed that, with a large probability all sufficiently fine networks (i.e., networks where each agent has sufficiently many neighbors), coordination on the special fuzzy convention of RU-dominant outcome is always an approximate equilibrium, regardless of the network structure.

The paper leaves many important questions unanswered. First, how the results extend to small-degree networks? Second, in real applications, both macro- and micro-level heterogeneity are observed. Likely, the latter is due to systematic differences in preferences (Perhaps, differences in the threshold distributions) across different parts of the network. Can the two idiosyncratic and systematic differences be combined in a single model? Third, and related, can real world data be used to estimate parameters of the model, like the threshold distribution function $P(.)$? We leave these questions for future research.

Appendix A. Proof of Theorem 1

A.1. **Proof of Lemma 1.** Define a distance on the space of (mixed) profiles: For any $a, b \in [0,1]^N$, let

$$d(a,b) = \sqrt{\frac{1}{\sum g_i^2} \sum g_i^2 (a_i - b_i)^2}.$$

Recall that $\mathcal{B} = \{\beta^a : a \text{ is action profile}\}$ is the space of neighborhood fractions. For each $\delta > 0$, let $\mathcal{N}(\delta, \mathcal{B})$ be the covering number of $\mathcal{B}$, i.e., the smallest cardinality n of a list of profiles $b^1, ..., b^n \in \mathcal{B}$ such that, for each $b \in \mathcal{B}$, there is $l \leq n$ so that $d\left(b, b^l\right) \leq \delta$.

Lemma 2. *There exists a universal constant $c < \infty$ such that, for each $\delta > 0$, and each network g,*

$$\mathcal{N}(\delta, \mathcal{B}) \leq \exp\left(\frac{1}{\delta^2} c w^{*2} d(g) N\right).$$

Proof. We will use Sudakov's Minoration Inequality (Theorem 7.4.1 from Vershynin (2018)), which provides an upper bound on the covering number via the expectation of a certain Gaussian process. For this, let Z_i for each agent i be an i.i.d. standard normal random variable. For each (possibly mixed) profile $a \in \mathcal{A}$, define

$$X_a = \frac{1}{\sqrt{\sum_i g_i^2}} \sum_i g_i a_i Z_i.$$

For any two profiles $a, b \in \mathcal{A}$,

$$\sqrt{\mathbb{E}(X_a - X_b)^2} = \sqrt{\frac{1}{\sum g_i^2} \mathbb{E}\left(\sum_i g_i (a_i - b_i) Z_i\right)^2}$$
$$= \sqrt{\frac{1}{\sum g_i^2} \sum_i g_i (a_i - b_i)^2} = d(a,b).$$

Given the definition and the above property, Sudakov's Minoration Inequality implies that, for some universal constant $c_1 > 0$ (i.e., a constant that is independent of parameters and the current problem),

$$\log \mathcal{N}(\delta, \mathcal{B}) \leq c_1 \frac{\left(\mathbb{E} \sup_{b \in \mathcal{B}} X_b\right)^2}{\delta^2}.$$

We compute

$$\begin{aligned}
\mathbb{E}\sup_{b\in\mathcal{B}} X_b = \mathbb{E}\sup_{a\in\mathcal{A}} X_{\beta^a} &= \mathbb{E}\left(\sup_{a\in\mathcal{A}} \frac{1}{\sqrt{\sum_i g_i^2}} \sum_i g_i Z_i \left(\frac{1}{g_i}\sum g_{ij} a_j\right)\right) \\
&= \frac{1}{\sqrt{\sum_i g_i^2}}\,\mathbb{E}\left(\sup_{a\in\mathcal{A}} \sum_i a_i \left(\sum_j g_{ij} Z_j\right)\right) \leq \frac{1}{\sqrt{\sum_i g_i^2}}\,\mathbb{E}\sum_i \left|\sum_j g_{ij} Z_j\right| \\
&\leq \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{\sum_i g_i^2}} \sum_i \sqrt{\sum_j g_{ij}^2},
\end{aligned}$$

where the last inequality is due to a bound on the expectation of the absolute value of the normal variable $\sum g_{ij} Z_j$ via its standard deviation $\sigma_i = \sqrt{\sum_j g_{ij}^2}$. Because $\sum_j g_{ij}^2 \leq d(g)\, g_i^2$ and $\left(\sum_i g_i\right)^2 \leq N^2 w^{*2} g_{\min}^2 \leq N w^{*2} \sum g_i^2$, we have

$$\log \mathcal{N}(\delta, \mathcal{B}) \leq \sqrt{\frac{2}{\pi}} c_1 \frac{1}{\delta^2} \frac{1}{\sum_i g_i^2} \left(\sum_i \sqrt{d(g)} g_i\right)^2 d(g) \leq \frac{1}{\delta^2}\sqrt{\frac{2}{\pi}} c_1 w^{*2} d(g)\, N.$$

□

We proceed with the proof of Lemma 1. For the first inequality, suppose f is K-Lipschitz. Fix $\varepsilon > 0$ and $\delta > 0$ so that $\delta = \frac{1}{12K\sqrt{w^*}}\varepsilon$. Find δ-cover $b^1, ..., b^n$ of $\mathcal{B}$. Because $n \leq \mathcal{N}(\delta, \mathcal{B})$, Lemma 2 implies that

$$\begin{aligned}
&\text{Prob}\left(\sup_{l\leq n}\left|\sum_i g_i f\left(\tau_i, b_i^l\right) - \sum_i g_i\,\mathbb{E}\, f\left(., b_i^l\right)\right| \geq \frac{1}{2}\varepsilon \sum g_i\right) \\
\leq & nB\exp\left(-c\left(\frac{1}{2}\varepsilon\right) N\right) \leq B\exp\left(-\left(c\left(\frac{1}{2}\varepsilon\right) - \frac{1}{\delta^2} c w^{*2} d(g)\right) N\right).
\end{aligned}$$

Assume that the complement of the event in the parentheses of the first line of the above inequality holds. For each action profile a, find l so that $d\left(b^l, \beta^a\right) \leq \delta$. Then, by the Jensen's inequality, and because $\frac{g_i}{\sum g_i} \leq w^* \frac{g_i^2}{\sum g_i^2}$,

$$\sum_i \frac{g_i}{\sum_i g_i}\left|\beta_i^a - b_i^l\right| \leq \sqrt{\sum \frac{g_i}{\sum g_i}\left(\beta_i^a - b_i^l\right)^2} \leq \sqrt{\sum w^* \frac{g_i^2}{\sum g_i^2}\left(\beta_i^a - b_i^l\right)^2} \leq \sqrt{w^*}\delta.$$

Hence,

$$\begin{aligned}
&\left|\sum_i g_i f\left(\tau_i, \beta_i^a\right) - \sum_i g_i \mathbb{E} f\left(., \beta_i^a\right)\right| \\
&\leq \left|\sum_i g_i f\left(\tau_i, b_i^l\right) - \sum_i g_i \mathbb{E} f\left(., b_i^l\right)\right| + 2K \left|\sum_i g_i \left|\beta_i^a - b_i^l\right|\right| \\
&\leq \left|\sum_i g_i f\left(\tau_i, b_i^l\right) - \sum_i g_i \mathbb{E} f\left(., b_i^l\right)\right| + 2K\sqrt{w^*}\delta \left(\sum_i g_i\right) \leq \varepsilon \sum_i g_i.
\end{aligned}$$

Take $c\left(\varepsilon, K, d\right) = c\left(\frac{1}{2}\varepsilon\right) - \frac{1}{\varepsilon^2} c\left(6Kw^*\right)^2 d$. The claim follows.

For the second inequality, we first derive a version of (7): we show that there exists constants $B < \infty$ and $c\left(\varepsilon\right) > 0$ such that, for each profile a and measurable function $f\left(\tau, \beta\right) \in [0, 1]$,

$$\text{Prob}\left(\left|\sum_i g_{ij}\left(\prod_{k=i,j} f\left(\tau_k, \beta_k^a\right)\right) - \sum_i g_{ij}\left(\prod_{k=i,j} \mathbb{E} f\left(., \beta_k^a\right)\right)\right| \geq \varepsilon \sum g_i\right) \leq B \exp\left(-c\left(\varepsilon\right) N\right). \tag{13}$$

Indeed, suppose that $X_i \in [-1, 1]$ is a collection of independent mean zero random variables. The Hanson-Wright inequality (Theorem 6.2.1 Vershynin (2018)) implies that there exists a universal constant $c > 0$ such that, for each $t > 0$,

$$\mathbb{P}\left(\left|\sum g_{ij} X_i X_j - \mathbb{E} \sum g_{ij} X_i X_j\right| \geq t\right) \leq 2 \exp\left(-c \min\left(\frac{t^2}{\|G\|_F^2}, \frac{t}{\|G\|}\right)\right). \tag{14}$$

where $G = [g_{ij}]$ is the adjacency matrix, $\|G\|_F$ is the Frobenius norm and $\|G\|$ is the operator L^2-norm. Let $X_i = f\left(\tau_i, \beta_i^a\right) - \mathbb{E} f\left(., \beta_i^a\right)$ and $t = \varepsilon_i \sum g_i$. Recall that X_i are independent and that $g_{ij} = g_{ji}$ and $g_{ii} = 0$ to obtain

$$\begin{aligned}
&\sum g_{ij} X_i X_j - \mathbb{E} \sum g_{ij} X_i X_j = \sum g_{ij} X_i X_j \\
&= \sum g_{ij}\left(\prod_{k=i,j} f\left(\tau_k, \beta_k^a\right)\right) - \sum_i g_{ij}\left(\prod_{k=i,j} \mathbb{E} f\left(., \beta_k^a\right)\right) \\
&\quad - 2 \sum g_i \left(\mathbb{E} f\left(., \beta_j^a\right)\right)\left(f\left(\tau_i, \beta_i^a\right) - \mathbb{E} f\left(., \beta_i^a\right)\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\operatorname{Prob}\left(\left|\sum_i g_{ij}\left(\prod_{k=i,j} f\left(\tau_k, \beta_k^a\right)\right)-\sum_i g_{ij}\left(\prod_{k=i,j} \mathbb{E} f\left(., \beta_k^a\right)\right)\right| \geq \varepsilon \sum g_i\right) \\
\leq &\operatorname{Prob}\left(\left|\sum g_{ij} X_i X_j-\mathbb{E} \sum g_{ij} X_i X_j\right| \geq \frac{1}{2} \varepsilon \sum g_i\right) \\
&+\operatorname{Prob}\left(\left|\sum g_i\left(\mathbb{E} f\left(., \beta_j^a\right)\right)\left(f\left(\tau_i, \beta_i^a\right)-\mathbb{E} f\left(., \beta_i^a\right)\right)\right| \geq \frac{1}{4} \varepsilon \sum g_i\right).
\end{aligned}$$

We apply (14) to the first bound (notice that $\|G\| \leq \|G\|_F \leq \sqrt{N}\,\|G\|$ and $g_{\min} \leq \|G\| \leq w^* g_{\min}$, where $g_{\min} = \min_i g_i$) and Hoeffding's inequality (7) to the second bound to obtain

$$\begin{aligned}
\leq &2 \exp\left(-c \min\left(\varepsilon^2 \frac{\left(\sum g_i\right)^2}{N w^{*2} g_{\min}^2}, \varepsilon \frac{\sum g_i}{w^* g_{\min}}\right)\right)+B \exp\left(-c\left(\frac{1}{4} \varepsilon\right) N\right) \\
\leq &2 \exp\left(-c \frac{1}{w^{*2}} \varepsilon^2 N\right)+B \exp\left(-c\left(\frac{1}{4} \varepsilon\right) N\right).
\end{aligned}$$

This concludes the proof of (13).

Given (13), we conclude the proof of the second inequality of Lemma 1 in the same manner as in the case of the first inequality. In particular, if $d\left(b^l, \beta^a\right) \leq \delta$,

$$\begin{aligned}
&\left|\sum g_{ij}\left(\prod_{k=i,j} f\left(\tau_k, \beta_k^a\right)-\prod_{k=i,j} f\left(a_k, b_k^l\right)\right)\right| \\
\leq &\sum g_{ij} f\left(\tau_i, \beta_i^a\right)\left|f\left(\tau_j, \beta_j^a\right)-f\left(\tau_k, b_j^l\right)\right|+\sum g_{ij} f\left(\tau_j, b_j^l\right)\left|f\left(\tau_i, \beta_i^a\right)-f\left(\tau_i, b_i^l\right)\right| \\
\leq &K\left(\sum_j\left(\sum_i g_{ij} f\left(\tau_i, \beta_i^a\right)\right)\left|\beta_j^a-b_j^l\right|+\sum_i\left(\sum_j g_{ij} f\left(\tau_j, b_j^l\right)\right)\left|\beta_i^a-b_i^l\right|\right) \\
\leq &2K \sum_i g_i\left|\beta_i^a-b_i^l\right| \leq 2K \sqrt{w^*} \delta \sum_i g_i \leq \frac{1}{2} \varepsilon \sum_i g_i.
\end{aligned}$$

Similar calculations apply to $\sum_i g_{ij}\left(\prod_{k=i,j} \mathbb{E} f\left(., \beta_k^a\right)\right)$. Hence, if

$$\sup_{l \leq n}\left|\sum_i g_{ij} \prod_{k=i,j} f\left(\tau_k, b_k^l\right)-\sum_i g_{ij} \mathbb{E} \prod_{k=i,j} f\left(., b_k^l\right)\right| \geq \frac{1}{2} \varepsilon \sum g_i, \text{ for each } l$$

then

$$\left|\sum_i g_{ij}\left(\prod_{k=i,j} f\left(\tau_k, \beta_k^a\right)\right) - \sum_i g_{ij}\left(\prod_{k=i,j} \mathbb{E} f\left(., \beta_k^a\right)\right)\right| \leq \varepsilon \sum_i g_i.$$

The rest of the argument follows.

A.2. **Proof of Theorem 1.** Fix $\eta > 0$. For each $\delta > 0$, let $\nu_\delta^0 = \max_{x:|x-x^*|\leq\delta}\left(\nu\left(x^*\right) - \nu\left(x\right)\right)$ and let $\nu_\delta^1 = \min_{x:|x-x^*|\geq\delta}\left(\nu\left(x^*\right) - \nu\left(x\right)\right)$. Because x^* is the unique maximizer of $\nu\left(.\right)$, $\nu_\delta^1 > 0$ for each δ. Moreover, $\lim_{\delta\to 0}\nu_\delta^0 > 0$.

Let $\kappa > 0$ and define $\frac{1}{\kappa}$-Lipshitz functions:

$$\begin{aligned}\mathbf{1}^-\left(\tau, \beta\right) &= \max\left(0, \min\left(1, \frac{1}{\kappa}\left(\beta - \tau\right)\right)\right),\\ \mathbf{1}^+\left(\tau, \beta\right) &= \max\left(0, \min\left(1, 1 + \frac{1}{\kappa}\left(\beta - \tau\right)\right)\right).\end{aligned}$$

Then, $\mathbf{1}\left(\tau \leq \beta - \kappa\right) \leq \mathbf{1}^-\left(\tau, \beta\right) \leq \mathbf{1}\left(\tau \leq \beta\right) \leq \mathbf{1}^+\left(\tau, \beta\right) \leq \mathbf{1}\left(\tau \leq \beta + \kappa\right)$.

For any any equilibrium profile $a_i = \mathbf{1}\left(\tau \leq \beta_i^a\right)$, we have

$$V\left(a; \tau\right) \leq \frac{1}{2}\sum g_{ij}\mathbf{1}^+\left(\tau_i, \beta_i^a\right)\mathbf{1}^+\left(\tau_j, \beta_j^a\right) - \sum g_i \mathbf{1}^-\left(\tau_i, \beta_i^a\right)\tau_i.$$

An application of probabilistic bounds (7) and (13) shows that, if N is sufficiently large, then, with a probability of at least $1-\varepsilon$,

$$\begin{aligned}V\left(a^*; \tau\right) &\geq \frac{1}{2}\sum g_{ij} P\left(x^* - \kappa\right) P\left(x^* - \kappa\right) - \sum g_i \int_0^{x^*+\kappa} y dP\left(y\right) - \varepsilon \sum g_i\\ &= \sum g_i\left(\nu\left(x^* - \kappa\right) - \varepsilon - 2\kappa\right)\\ &\geq \sum g_i\left(\nu\left(x^*\right)\right) - \kappa N w^* g_{\min}\nu_\kappa^0 - \left(\varepsilon + 2\kappa\right) N w^* g_{\min},\end{aligned}$$

where in the last inequality, we use constants $\nu_.^0$.

Because $\mathbb{E}\,\mathbf{1}^-\left(., b\right) \geq P\left(b - \kappa\right)$ and $\mathbb{E}\,\mathbf{1}^+\left(., b\right) \leq P\left(b + \kappa\right)$, an application of Lemma 1 shows that, for each $\varepsilon > 0$, there is $d > 0$ small enough such that if $d\left(g\right) < d$ (hence N is sufficiently large), then with probability of at least $1-\varepsilon$, we have for each equilibrium profile a,

$$\begin{aligned}V(a;\tau) \leq & \frac{1}{2}\sum g_{ij}P\left(\beta_i^a+\kappa\right)P\left(\beta_j^a+\kappa\right)-\sum g_i\int_0^{b_i-\kappa} ydP(y)+\varepsilon\sum g_i \\ \leq & \frac{1}{2}\sum g_i\left(P\left(\beta_i^a+\kappa\right)\right)^2-\sum g_i\int_0^{b_i-\kappa} ydP(y)+\varepsilon\sum g_i \\ = & \sum g_i\left(\nu\left(\beta_i^a+\kappa\right)+\varepsilon+2\kappa\right).\end{aligned}$$

If an equilibrium profile $a=\mathbf{1}\left(\tau_i \leq \beta_i^a\right)$ is not an η-fuzzy convention, then we get

$$V(a;\tau) \leq \sum g_i\left(\nu\left(x^*\right)\right)+Nw^*g_{\min}(\varepsilon+2\kappa)-\eta Ng_{\min}\nu^1_{\frac{1}{2}\eta}$$

,where we used the definition of constants $\nu^1_{\cdot}$. If κ and $\varepsilon \leq \frac{1}{2}\eta$ (and $d(g)$) are sufficiently small, $V(a;\tau)<V\left(a^*;\tau\right)$ with probability of at least $1-2\varepsilon \geq 1-\eta$. In such a case, the potential maximizer must be an η-fuzzy convention x^*.

Appendix B. Proof of Theorem 2

In part B.1 of this Appendix, we formally define the city network (M,m) and also develop some of its properties. Part B.2 contains the probabilistic part of the proof: We establish the existence of a large connected component of the network that is also obstacle-free, i.e., without "bad" groups of agents. The last part elaborates on the contagion argument from the main body of the paper to conclude the proof of the Theorem.

B.1. **Lattice.** We start by formally defining the city network. For each $M \geq m$, the (M,m)-lattice is a network with

- $N=M^2$ nodes from the set $I_M=\{1,\ldots,M\}^2$. We define a distance on I_M by
$$d(i,j)=\frac{1}{m}\sqrt{\sum_l\left(\left(i_l-j_l\right) \mod M\right)^2},$$
and a ball in this metric as $B(i,r)=\{y: d(x,y) \leq r\}$. The subtraction "mod$M$" turns the lattice into a subset of "discrete Euclidean torus" $\left[0,\frac{M}{m}\right]^2$,
- connections $g_{i,j}=1 \Longleftrightarrow j \in B(i,1)$.

For each $i \in I_M$, and two sets $U, W \subseteq I_M$, let

$$d(i, W) = \min_{j \in W} d(i, j) \text{ and } d(U, W) = \min_{i \in U} \min_{j \in W} d(i, j). \tag{15}$$

For each set W, and each r, define the r-neighborhood of W:

$$B(W, r) = \{i : d(i, W) \leq r\} = \bigcup_{i \in W} B(i, r).$$

B.1.1. *Large m approximations.* For large m, the neighborhoods of each agent have similar properties as open balls on a Euclidean plane. This is formalized as follows. Let $B_{\mathbb{R}^2}(x, r)$ be the ball on the plane with center $x \in \mathbb{R}^2$ and radius r. Let $|A|$ be a Lebesgue measure of a measurable set $A \subseteq \mathbb{R}^2$. Let

$$f_0(d, r_1, r_2) = \frac{1}{\pi} |B_{\mathbb{R}^2}((0,0), r_1) \cap B_{\mathbb{R}^2}((d,0), r_2)|$$

be the mass of the intersection of two balls, with radii r_1 and r_2 respectively, separated by distance d, and normalized by the mass of the unit ball $B((0,0), 1)$.

Lemma 3. *The following holds:*

- *For each $\rho > 0$, there exists $C_\rho < \infty$ such that if $m \geq C_\rho$, then for any two agents i, j, for any $r_1 \leq 1 \leq r_2$, we have*
$$\left| \frac{|B(i, r_1) \cap B(j, r_2)|}{|B(i, 1)|} - f_0(d(i, j), r_1, r_2) \right| \leq \rho.$$
- *Function f_0 has the following properties:*
  - *f_0 is Lipschitz over d and $r_1 \leq 1 \leq r_2$,*
  - *f_0 is decreasing in d, and*
  - *$f_0(d, r_1, r_2) = 0$ if $r_1 + r_2 \leq d$, and $f_0(d, r_1, r_2) = 1$ if $r_1 = 1$ and $d \leq r_2 - r_1$.*
- *Functions $f_1(x, r_1; r_2) = f_0(r_2 - x, r_1, r_2)$ for $r_1 \leq 1$ and $x \in \mathbb{R}$ converge uniformly to function $\lim_{r_2 \to \infty} f_1(x, r_1; r_2) = f_2(x, r_1)$. In particular, for each $\rho > 0$, there exists R_ρ such that, if $r_1 \leq 1$ and $r_2 \geq R_\rho$, then,*
$$\sup_{r_1 \leq 1, x} |f_2(x, r_1) - f_1(x, r_1; r_2)| \leq \rho.$$
Functions f_1 and f_2 are Lipschitz over d and $r_1 \leq 1$ and increasing in x.
- *Let $f(x) = f_2(x, 1)$. Then, $f(x) + f(-x) = 1$.*

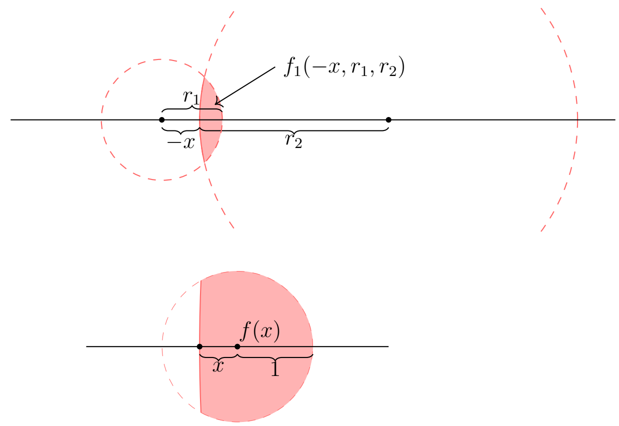


FIGURE 7. Illustrations of functions f_1 and f.

Proof. The properties of f_0, f_1, f_2, and f follow from their geometric interpretations and from the fact that the counting measure on I_M converges weakly to the Lebesgue measure on the torus. For example, $f_2(x, r_1)$ is a segment of radius r_1 ball with height equal to $r_1 + x$ for $x \in (-r_1, r_1)$. See Figure 7. □

B.1.2. *Cubes.* let G be a (M, m)-lattice. We divide the lattice into disjoint areas that we refer to as *cubes*. We will assume there exists values b such that $0 \ll b \ll m$, and M is divisible by b. (This divisibility assumption simplifies the proof. The theorem remains valid without it, but the proof requires small modifications due to the existence of non-zero reminders from the division by b. We omit the details.) Each cube has b^2 elements and, because $b \ll m$, it is much smaller than the diameter of the neighborhood of each node so that the neighborhoods of nodes in the same cube are largely overlapping. At the same time, each cube contains a sufficiently large number of nodes so that the distribution of thresholds within the cube can be probabilistic ally approximated by its expected distribution.

Formally, for each real number x, let $\lfloor x \rfloor$ be the largest integer no larger than x. For each node i, the set of nodes

$$c^b(i) = \left\{ j \in \{1, ..., M\}^2 : \forall_l \lfloor i_l/b \rfloor = \lfloor j_l/b \rfloor \right\}$$

is referred to as a cube that contains i. Any two cubes are either disjoint or identical. Each cube c is uniquely identified by a pair of numbers $c_l = \lfloor i_l/b \rfloor$ for each $l = 1, 2$ and any $i \in c$. Due to the divisibility assumption, there are $\left(\frac{M}{b}\right)^2$ cubes on the (M, m)-lattice.

Let $\mathcal{G}^b = \left\{c^b(i) : i \in G\right\}$ be the set of all cubes. We refer to the elements of $\mathcal{G}^b$ as *cubes*. The network of cubes $\mathcal{G}^b$ consists of cubes as vertices and edges between any two cubes that share one of their sides: for any $c, c' \in \mathcal{G}^b$, $g^b_{c,c'} = 1$ iff for some $l = 1, 2$, $c_l = c'_l$ and $\left|\left(c_{-l} - c'_{-l}\right) \mod \frac{M}{b}\right| = 1$. Thus, each cube shares an edge with four other cubes.

Say that set $S \subseteq \mathcal{G}^b$ is r-*connected* if for any subset $A \subseteq S, A \neq S$, there is $c \in A$, $c' \in S \backslash A$, and at most an r-element path between c and c'. (A path is a tuple of cubes connected by the edges of the cube network.) S is connected if it is 1-connected.

For any two cubes, define a distance $d^b(c, c') = \max_l \left|(c_l - c'_l) \bmod \frac{M}{b}\right|$. For any $S, S' \subseteq \mathcal{G}^b$, let $d^b(S, S') = \min_{c \in S, c' \in S'} d^b(c, c')$ be the distance between two sets of cubes. Let $U(c, r) = \{c' : d(c, c') \leq r\}$ be the r-neighborhood of c. Thus, each cube has 8 other cubes in its 1-neighborhood.

B.2. **Probabilistic part.** We will show that if the lattice is sufficiently large then, with arbitrarily high probability, we can find a set W of cubes that (a) contains almost all cubes and (b) is connected in the cube network, where (c) each cube in the set is far away from bad cubes, and (d) contains a large set of agents for whom action 0 is dominant. Properties (b)-(c) will allow the contagion wave to spread across the entire set W, property (a) will ensure that spreading to set W means spreading almost everywhere, and property (d) will ensure that the set contains sufficiently many "initial infectors" to start the contagion wave.

For each realization of threshold profile τ, define the empirical cdf of best response thresholds in cube $c \in \mathcal{G}^b$:

$$P_c(x|\tau) = \frac{1}{|c|} \sum_{i \in c} \mathbf{1}\{\tau_i < x\}.$$

For $\gamma > 0$, say that a cube c is γ-bad if there exists x such that $P_c(x|\tau) > P(x) + \gamma$; otherwise, the cube is γ-good.

Agent x is *extraordinary* if action 0 is strictly dominant for such an agent. A cube $c \in \mathcal{G}^b$ is *extraordinary* if it only consists of extraordinary agents. In any equilibrium,

$a(c) = 0$ for extraordinary cube c. Clearly, an extraordinary cube is γ-good for each $\gamma \geq 0$.

Say that set $W \subseteq \mathcal{G}^b$ of cubes is (γ, R)-*good* if

(a) W contains at least a fraction $(1-\gamma)$ of cubes, $|W| \geq (1-\gamma)\left|\mathcal{G}^b\right|$,
(b) W is connected as a subset of the cube network,
(c) if $c \in \mathcal{G}^b$ is γ-bad, then $d^b(c, c') > 3R$ for each $c' \in W$ (in particular, each cube in W is γ-good), and
(d) W contains a cube c_0 such that each cube c s.t. $d(c, c_0) \leq R$ is extraordinary.

We show that large good sets of cubes exist with high probability:

Lemma 4. *For each $\gamma, \rho > 0$, and $R < \infty$, there exist $m_{\gamma,\rho,R} > 0$, and for each $m > m_{\gamma,\rho,R}$, there exists $M_{\gamma,\rho,R}(m)$ such that, if $m \geq m_{\gamma,\rho,R}$ and $M \geq M_{\gamma,\rho,R}(m)$, then, if G is an (M,m)-lattice, $b = \lfloor \rho m \rfloor$, and $\mathcal{G}^b$ is the associated cube network, then*

$$\mathbb{P}\left(\textit{there exists } (\gamma, R)\textit{-good set } W \subseteq \mathcal{G}^b\right) \geq 1-\gamma.$$

B.2.1. *Intermediate results.* We need two intermediate results. The first result provides a bound on the size of the largest connected component of the graph obtained from the network of cubes after removing a group of smaller and connected sets of cubes.

Lemma 5. *Suppose that $\{S_1, ..., S_J\}$ is a collection of connected subsets of $\mathcal{G}^b$ such that $S_i \cup S_j$ are not 2-connected for any $i \neq j$. Then, there is a connected subset $V \subseteq \mathcal{G}^b \backslash \bigcup S_j$ such that $\left|\mathcal{G}^b \backslash V\right| \leq \sum_j |S_j|^2$.*

Proof. First, observe that for each connected set S such that $|S|^2 < \left|\mathcal{G}^b\right|$, there is a set S' and a loop (i.e., a path with the same beginning and ending) $c_0^S, ..., c_n^S = c_0$ of cubes $c_l^S \notin S'$ such that

- $S' \supseteq S$ and$|S'| \leq |S|^2$, and
- loop $c_0^S, ..., c_n^S$ tightly surrounds set S' and separates it from the rest of the graph: $|\{c : d(c, S') = 1\}| \subseteq \left\{c_l^S\right\} \subseteq |\{c : d(c, S') \leq 2\}|$.

This observation follows from the Jordan Curve Theorem and from the fact that each connected set S such that $|S|^2 < \left|\mathcal{G}^b\right|$ can be contained in a $|S'^2|$-element "square" of cubes such that the set outside the square is connected.

For each set S_i from the hypothesis of the Lemma, find loop c^i and set S_i' as in the observation above. We will show that set $\mathcal{G}^b \backslash \bigcup S_j'$ is connected, which will conclude the proof of the Lemma. Take any two cubes $c, c' \in \mathcal{G}^b \backslash \bigcup S_j'$ and an arbitrary path $c = c_0, ..., c_n = c'$ between them. We will modify this path so that it avoids each set S_i. For each i, either the existing path avoids set S_i', or it intersects it. Find $l_0^i = \min \{l : d(c_l, S_i) = 1\}$ and $l_1^i = \max \{l : d(c_l, S_i) = 1\}$. Then, replace the interval $c_{l_0^i}, ..., c_{l_1^i}$ of the path with the path from $c_{l_0^i}$ to $c_{l_0^i}$ along path c^i. The new path avoids set S_i'. Because the modified part of the path stays within 2-distance of set S_i', the modification does not create new intersections with other sets S_j'. After possibly modifying the path for any i, we obtain a path between c and c' that avoids each set S_i'. Thus, set $\mathcal{G}^b \backslash \bigcup S_j'$ is connected. □

The second result provides an upper bound on the number of different r-connected sets of cubes.

Lemma 6. *The number of r-connected sets in $\mathcal{G}_b$ of cardinality n is not larger than $2^{2n} (2r+1)^n |\mathcal{G}_b|$.*

Proof. We first find an encoding for each r-connected tuple. Let m_r be the size of the r-neighborhood of an element of $\mathcal{G}_b$. Then, $m_r \leq (2r+1)^2$. Consider tuples $(s_1, (l_2, ..., l_n), (k_2, ..., k_n))$ such that $s_1 \in \mathcal{G}_b$, $k_i \in \{1, .., m_r\}$, and $l_i \leq i$ and $l_i \leq l_j$ for each $2 \leq i \leq j$.

We show that each r-connected set can be encoded as one of the above tuples in such a way that any two different r-connected sets must have a different encoding. Let $e : \mathcal{G}^b \to \left\{1, ..., \left|\mathcal{G}^b\right|\right\}$ be an enumeration of set $\mathcal{G}^b$. For each $s \in \mathcal{G}^{\lfloor}$,let $e_s : \{s' : d(s, s') = 1\} \to \{1, ..., 4\}$ be the enumeration of the immediate neighborhood of s that has the same ranking in the neighborhood as enumeration e. Choose $s_1 = \arg\min_{s \in S} e(s)$. Suppose that $s_1, ...s_{i-1}$ are chosen for $1 < i < n$. For each $x \in S \backslash \{s_1, ..., s_{i-1}\}$, let $l(x) = \min_{d(x, s_l)=1} l$ and let it equal ∞ if the set is empty. Then, $l(x) < i$ for at least one x. Let $k(x) = e_{s_{l(x)}}(x)$. Choose $s_i = \arg\min_{\text{lexicograpically}, x \in S} (l(x), k(x))$, so as to minimize lexicographically $(l(x), k(x))$ among all $x \in S \backslash \{s_1, ..., s_{i-1}\}$. Let $l_i = l(s_i)$ and $k_i = k(s_i)$.

We derive an upper bound on the number of encoding tuples. Say that a sequence $l_i, ..., l_n$ is (i, m)-sequence if it is increasing, $l_j < j$ for each j, and $l_i = i - m - 1$. Let

$S(i,m)$ denote the number of different (i,m)-sequences. It is easy to see that

$$S(i,m) = \sum_{p=0}^{m+1} S(i+1,p),$$

where $S(n,m)=1$. We check by induction on i that $S(i,n) \leq 2^{2(n-i)+m}$.

The number of choices for s_1 is not larger than $|\mathcal{G}_b|$. By the above, the number of $(2,0)$-sequences is not larger than $2^{2(n-2)}$. The number of choices of $k_2,...,k_n$ is not larger than $(2r+1)^{n-1}$. It follows that the total number of encodings, and hence the number of connected sets, is not larger than $2^{2n}(2r+1)^n|\mathcal{G}_b|$. $\square$

B.2.2. *Proof of Lemma 4.* Lemma 4 follows from the following two results. The first result establishes the existence of a large connected component that is far from bad cubes. Let $B_\gamma = \left\{c \in \mathcal{G}^b : c \text{ is } \gamma\text{-bad}\right\}$ be the (random) set of γ-bad cubes.

Lemma 7. *For each $\gamma > 0$ and $R < \infty$, there exists $b_{\gamma,R} > 0$ such that if $b > b_{\gamma,R}$, then*

$$\mathbb{P}\left(\exists W^0 \subseteq \mathcal{G}^b, \textit{ st. } W^0 \textit{ is connected, } \left|W^0\right| \geq (1-\gamma)\left|\mathcal{G}^b\right|,\ d^b\left(W^0, B_\gamma\right) \geq 5R\right) \geq 1-\frac{1}{4}\gamma.$$

Proof. Let $p_\gamma > 0$ be the probability that a cube is γ-bad. Due to the Dvoretzky–Kiefer–Wolfowitz–Massart inequality, the probability that a cube c is γ-bad is bounded by

$$p_\gamma \leq C\mathrm{e}^{-2b^2\gamma^2}$$

for some universal constant C.

Let $S_1^0,...,S_n^0$ be the smallest division of the set of bad cubes $B_\gamma = \bigcup S_i^0$ into sets that are $11R$-connected and such that $S_i^0 \cup S_j^0$ are not $11R$-connected for $i \neq j$. Let $X = \sum |S_i^0|^2$. We compute the expected value of X. Let $m_n = (2^{2n}(11R+1)^n|\mathcal{G}_b|)$ be an upper bound on the cardinality of all $11R$-connected sets (obtained from Lemma 6). Then,

$$\begin{aligned}\mathbb{E}\,X &\leq \sum_{n\geq 1} n^2 m_n p_\gamma^n \leq |\mathcal{G}_b| \sum_{n\geq 1} 2^n 2^{2n}(6R+1)^n p_\gamma^n \\ &= |\mathcal{G}_b| \frac{8(11R+1)p_\gamma}{1-8(11R+1)p_\gamma}.\end{aligned}$$

Let $S_i^1 \supseteq S_i^0$ be the smallest connected set such that sets $S_i^1 \cup S_j^1$ are not $11R$-connected for $i \neq j$ and such that $|S_i^1| \leq 11R\,|S_i^0|$. Such sets can be constructed by connecting elements of S_i^0 by a path inside the intersection of the $11R$-neighborhood of the two sets.

Let S_i be the $5R$-neighborhood of set S_i^1. Clearly, sets S_i are disjoint (and separated by R). Because each $5R$-neighborhood of an element of a set S_i^1 has no more than $(11R+1)^2\,|S_i^1|$ cubes, the cardinality of S_i is at most $(11R+1)^2\,|S_i^1| \leq (11R+1)^3\,|S_i^0|$.

Let W^0 be the largest connected component of $\mathcal{G}^b$ that does not contain elements of sets S_i. By construction, each set S_i is connected, but sets $S_i \cup S_j$ are not 2-connected. By Lemma 5, the cardinality of W^0 is at least $\left|\mathcal{G}^b\right| - 4\,(11R+1)^6\,X$. By the Markov's inequality,

$$\begin{aligned}\mathbb{P}\left(\left|W^0\right| \geq (1-\gamma)\left|\mathcal{G}^b\right|\right) &\leq \mathbb{P}\left(4\,(11R+1)^6\,X \leq \gamma\left|\mathcal{G}^b\right|\right)\\ &\leq \frac{4\,(11R+1)^6\,\mathbb{E}\,X}{\gamma\,|\mathcal{G}^b|} \leq \frac{1}{\gamma}\frac{32\,(11R+1)^7\,p_\gamma}{1-8\,(11R+1)\,p_\gamma}.\end{aligned}$$

Assume that $b_{\gamma,R} > 0$ is large enough so that for each $b > b_{\gamma,R}$, $\frac{1}{\gamma}\frac{32(11R+1)^7 C\mathrm{e}^{-2b^2\gamma^2}}{1-8(11R+1)C\mathrm{e}^{-2b^2\gamma^2}} \leq \frac{1}{4}\gamma$. □

Say that cube $c \in G_R$ is an extraordinary center if all cubes in $U(c,R)$ are extraordinary.

Lemma 8. *There exists $K_{\gamma,R} < 0$ large enough so that if $\frac{M}{b} > K_{\gamma,R}$, then*

$$\mathbb{P}\left(\exists W \subseteq \mathcal{G}^b,\ \textit{st.}\ \begin{array}{c} W \supseteq W^0,\ W \textit{ is connected, } d^b(W,B_\gamma) \geq 3R \\ \textit{and } W \textit{ contains an extraordinary center}\end{array}\right) \geq 1-\gamma,$$

where W_0 inside the probability satisfied the conditions from Lemma 7.

Proof. Recall that $K = \frac{M}{b}$ is the number of cubes. If K is divisible by $(2R+1)$, we can find a grid of cubes $G_R \subseteq \mathcal{G}^b$ such that any two $c, c' \in G$, $d(c,c') = 2R$ and $\mathcal{G}^b = \bigcup_{c\in G_R} U(c,R)$. Because the $U(c,R)$ neighborhoods are disjoint, $\left|\mathcal{G}^b\right| = |G_R|\,(2R+1)^2$, where $(2R+1)^2$ is the size of each neighborhood. For simplicity, the rest of the arguments rely on the divisibility assumption. The argument is easily modified for the case when the divisibility does not hold (and b and $\frac{M}{b}$ are sufficiently large).

Let W^0 be the (random) set from Lemma 7. Let $W^1 = \bigcup_c U(c, R+1)$ and $W = \bigcup_c U(c, 2R+1)$. Then, $d(W, B_\gamma) > 2R$. Because for each $c' \in U(c, r)$ there is a path between c and c' that is inside set $U(c, r)$, W is connected.

We show that $|G_R \cap W^1| \geq (1-\gamma)|G_R|$. On the contrary, suppose that $|G_R \backslash W^1| > \gamma \ |G_R|$. Then, $A = \bigcup_{c \in G_R \backslash W} U(c, R) \subseteq \mathcal{G}^b \backslash W^0$. Moreover, $|A| > \gamma |G_R| (2R+1)^2 = \gamma \left|\mathcal{G}^b\right|$. However, this contradicts $\left|\mathcal{G}^b \backslash W^0\right| \leq \gamma \left|\mathcal{G}^b\right|$.

Let $q > 0$ be the probability that a cube c is an extraordinary center. Then, $q \geq P(0)^{(2R+1)^2 b^2}$. Let q^* be the probability that cube c is an extraordinary center, conditional on $c \in W^1$. Because being in $c \in W^1$ provides no other information about the distribution of taste shocks apart from c is not γ-bad and γ-bad cubes are not extraordinary, it must be that $q^* \geq q$. Similarly, conditional on $c, c' \in W^1$, if c and c' are separated by $2R+1$, the events that the two are extraordinary centers are independent. Hence, the probability that none of the cubes in $c \in G_R \cap W_1$ is an extraordinary center is at most

$$\begin{aligned}(1-q^*)^{\left|G_R \cap W^1\right|} &\leq \left(1 - P(0)^{(2R+1)^2 b^2}\right)^{(1-\gamma)K^2(2R+1)^{-2}} \\ &\leq \mathrm{e}^{-(1-\gamma)K_{\gamma,R}(2R+1)^{-2} P(0)^{(2R+1)^2 b^2}}.\end{aligned}$$

If K is sufficiently large, the above is smaller than $\frac{1}{4}\gamma$. □

To conclude the proof of the Lemma, we set $m_{\gamma,\rho,R} > \frac{1}{\rho} b_{\gamma,R}$ and then $M_{\gamma,\rho,R}(m) \geq \rho m K_{\gamma,R}$.

B.3. **Proof of Theorem 2.** Below, we will show the following Lemma.

Lemma 9. *For each $\varepsilon > 0$, there exists sufficiently small $\gamma, \rho > 0$ and sufficiently large $R > 0$ so that if $b = \lfloor \rho m \rfloor$, W is a (γ, R)-good set in the network of cubes $\mathcal{G}^b$, and a is an equilibrium profile, then for each $i \in c \in W$, $\beta_i^a \leq x^* + \varepsilon$.*

Together with Lemma 4, Lemma 9 shows that for each $\varepsilon > 0$, if m and $\frac{M}{m}$ are sufficiently large, with probability of at least $1 - \varepsilon$, if a is an equilibrium profile, then all but a ε-fraction of the population (i.e., all members of the "good" set W), $\beta_i^a \leq x^* + \varepsilon$.

A similar argument shows that $\beta_i^a \geq x^* - \varepsilon$ for elements of an analogously defined "good" set (with the appropriate modification of what good and extraordinary cubes are). Together, the two arguments show that, with probability of at least $1-2\varepsilon$, each agent average neighborhood behavior is within ε of x^*. All such agents, if they have a threshold outside interval $[x^*-\varepsilon, x^*+\varepsilon]$, will choose best response as in 0-fuzzy convention x^* profile a^{x^*}.

Finally, choose ε small enough so that $P(x^*+\varepsilon) - P(x^*-\varepsilon) \leq \frac{1}{4}\eta$. Then, if the network is sufficiently large, the probability that the fraction of agents with threshold $\tau_i \in [x^*-\varepsilon, x^*+\varepsilon]$ is larger than $\frac{1}{2}\eta$ is smaller than ε.

Take $\varepsilon = \frac{1}{4}\eta$. Then, with a probability of at least $1-\eta$, at most $\frac{\eta}{2}$ agents have thresholds in the $\varepsilon-$interval, and at most $2\varepsilon = \frac{\eta}{2}$ agents observe equilibrium neighborhood behavior that is outside the $\varepsilon-$interval. All the other agents choose the same behavior as in profile a^{x^*}.

Proof. We divide the proof of the Lemma into two steps.

Preparation. Find $\varepsilon_0 > 0$, such that

$$\sigma^* = \max_{a \geq x^* + \frac{\varepsilon}{2}} \int\limits_{x^*+\varepsilon_0}^{a} \left(P^{-1}(y) - y\right) dy > 0.$$

The existence of such $\varepsilon_0 \in \left(0, \frac{\varepsilon}{2}\right)$ comes from the definition of x^* as the unique maximizer of $\int_{x^*}^{a} \left(y - P^{-1}(y)\right) dy$. Let δ_ρ be a fraction of neighbors of i who are not members of a cube that is fully contained in the neighborhood of i. It is easy to see that $\delta_\rho \to 0$ as $\rho \to 0$.

Let a be an equilibrium profile. For each cube c, define

$$a_c = \frac{1}{|c|} \sum_{j \in c} a_j \text{ and } \beta_c = \frac{1}{|c|} \sum_{j \in c} \beta_j^a.$$

Then, $|\beta_c - \beta_i^a| \leq \delta_\rho$, and

$$\beta_c \leq \delta_\rho + \frac{|c|}{|B(i,1)|} \sum_{c \subseteq B(i,1)} a_c. \tag{16}$$

If cube c is γ-good, then

$$a_c = \frac{1}{|c|} \sum_{i \in c} \mathbf{1}\left\{\tau_i < \beta_i^a\right\} \leq \frac{1}{|c|} \sum_{i \in c} \mathbf{1}\left\{\tau_i < \beta_c + \delta_\rho\right\} \leq P(\beta_c + \delta_\rho) + \gamma. \tag{17}$$

From now on, assume that $W \subseteq \mathcal{G}^b$ is (γ, R)-good. If $d^b(c, W) \leq 3R$, then cube c is γ-good.

Define

$$C_0 = \{c : \forall c' \; d(c, c') \leq R \Longrightarrow a_c \leq x^* + \varepsilon_0\}.$$

For each $i \in C_0$, the average behavior in all the cubes fully contained in the neighborhood of i is $\leq x^* + \varepsilon_0$, which, together with (16), implies that

$$\beta_i^a \leq (x^* + \varepsilon_0)(1 - \delta_\rho) + \delta_\rho \leq x^* + \varepsilon.$$

The last inequality holds when ρ is sufficiently small so that $\delta_\rho \leq \frac{\varepsilon}{2}$. Hence, to establish our claim, it is enough to show that $W \subseteq C_0$.

Notice that C_0 cannot be empty as it contains at least one extraordinary cube. For each $a > x^* + \frac{\varepsilon}{2}$, define

$$d(a) = \min_{c \in W : a_c \geq a} d^b(c, C_0) \geq R,$$

where the value is ∞ if the set over which the distance is minimized is empty.

On the contrary to our claim, suppose that there is a cube $c \in W_0$ such that $a_c > a > x^* + \frac{\varepsilon}{2}$. Then, there exists $a > x^* + \frac{\varepsilon}{2}$ such that $d(a) < \infty$. Find $a^* \geq x^* + \varepsilon_0$ such that $d(a^*) \leq 2R$ and $d\left(a^* + \frac{1}{R}\right) \geq d(a^*) + 1$. Such a^* exists: otherwise, if for each a such that $d(a) \leq 2R$, $d\left(a + \frac{1}{R}\right) \leq d(a) + 1$, then $d(a + 1) \leq 2R$, which is impossible (as there is no cube with the action average strictly larger than 1).

Contagion wave. Notice that a_c takes discrete values $a \in A = \left\{0, \frac{1}{|c|}, ..., 1\right\}$, where $|c|$ is the size of a cube. Let $a_k = \frac{k}{|c|}$ be the enumeration of set $A \cap \left\{a : a \geq x^* + \frac{\varepsilon}{2}\right\}$. For each such cube c, and each $i \in c$, (16) implies

$$\begin{aligned}
\beta_c \leq & \delta_\rho + \frac{|c|}{|B(i,1)|} \sum_{c \subseteq B(i,1)} a_c \\
\leq & \delta_\rho + \sum_{a \in A} a \frac{|\{c \subseteq B(i,1) : a_c = a\}|}{|B(i,1)| / c} \\
\leq & \delta_\rho + x^* + \varepsilon_0 + \sum_k (a_{k+1} - a_k) \frac{|\{c \subseteq B(i,1) : a_c \geq a\}|}{|B(i,1)| / c} \\
\leq & \delta_\rho + \delta_{R,\rho} + x^* + \varepsilon_0 + \sum_k (a_{k+1} - a_k) \left(1 - f\left(d(a_k) - d^b(c, C_0)\right)\right),
\end{aligned}$$

where the third inequality is a consequence of a discrete version of the integration by parts (i.e., $\sum x_i (y_i - y_{i+1}) = \sum (x_{i+1} - x_i) y_{i+1}$), and the fourth one is due to Lemma 3, where $\delta_{R,\rho} \to 0$ as R is sufficiently large and ρ is sufficiently small. Let $\delta^1_{R,\rho} = \delta_\rho + \delta_{R,\rho}$.

Additionally, for each $a_l \in A$, $a_l \leq a^*$, find a cube c such that $d^b(c, C_0) = d_R(a_l) < 2R$ and $a_c \geq a_l$. Using the above inequality and (17), we obtain

$$\begin{aligned} &P^{-1}(a_l - \gamma) \leq P^{-1}(a_c - \gamma) \leq \beta_c + \delta_\rho \\ \leq &\delta^1_{R,\rho} + x^* + \varepsilon_0 + \sum_k (a_{k+1} - a_k)(1 - f(d(a_k) - d(a_l))). \end{aligned}$$

Let $k^* = \max\{k : a_k \leq a^*\}$. Then, the right-hand side is not larger than

$$\begin{aligned} \leq &\delta^1_{R,\rho} + x^* + \varepsilon_0 + \sum_{k \leq k^*} (a_{k+1} - a_k)(1 - f(d(a_k) - d(a_l))) \\ &+ \sum_{k > k^* : a_k \leq a^* + \frac{\varepsilon}{10}} (a_{k+1} - a_k)(1 - f(d(a_k) - d(a_l))) \\ &+ \sum_{k : a_k > a^* + \frac{\varepsilon}{10}} (a_{k+1} - a_k)(1 - f(d(a_k) - d(a_l))) \\ \leq &\delta^1_{R,\rho} + x^* + \varepsilon_0 + \frac{1}{R} + \sum_{k \leq k^*} (a_{k+1} - a_k)(1 - f(d(a_k) - d(a_l))), \end{aligned}$$

due to the second term in the first line being not larger than $\frac{\varepsilon}{10}$, and the third term being equal to 0 (as $f(d(a_k) - d(a_l)) \geq f(1) = 1$).

Let $\Delta = a^* - (x^* + \varepsilon_0)$. Multiplying by $(a_{l+1} - a_l)$ and summing across $l \leq k^*$, we obtain

$$\begin{aligned} &\sum_{l \leq K^*} P^{-1}(a_l - \gamma)(a_{l+1} - a_l) \\ \leq &\left(\delta^1_{R,\rho + \frac{1}{R}} + x^*\right)\Delta + \sum_{l \leq K^*} \sum_{k \leq K^*} (a_{k+1} - a_k)(a_{l+1} - a_l)(1 - f(d_R(a_l) - d_R(a_k))) \\ = &\left(\delta^1_{R,\rho} + \frac{1}{R} + x^*\right)\Delta + \frac{1}{2} \sum_{l,k \leq K^*} (a_{k+1} - a_k)(a_{l+1} - a_l) = \left(\delta^1_{R,\rho} + \frac{1}{R} + x^* + \varepsilon_0\right)\Delta + \frac{1}{2}\Delta^2 \\ \leq &\delta^1_{R,\rho} + \frac{1}{R} + \int_{x^* + \varepsilon_0}^{a^*} y dy. \end{aligned}$$

To obtain the equality, we use the fact that f is balanced.

Because $P^{-1}(.-\gamma) \in [0,1]$ and $a_{l+1} - a_l = \frac{1}{|c|}$, the left-hand side of the above inequality is smaller than

$$\int_{x^*+\varepsilon_0}^{a^*} P^{-1}\left(y-\gamma-\frac{1}{|c|}\right) dy \geq \int_{x^*+\varepsilon_0-\gamma-\frac{1}{|c|}}^{a^*-\gamma-\frac{1}{|c|}} P^{-1}(y)\, dy$$

Assuming that b is large enough so that $\frac{1}{|c|} \leq \gamma$, the above is not smaller than $\int_{x^*+\varepsilon_0}^{a^*} \left(P^{-1}(y) - y\right) dy - 2\gamma$. Putting it back into the main inequality, we obtain

$$\int_{x^*+\varepsilon_0}^{a^*} \left(P^{-1}(y) - y\right) dy \leq \delta^1_{R,\rho} + \frac{1}{R} + 2\gamma.$$

If $\gamma, \rho > 0$ are sufficiently small and R sufficiently large, $\delta^1_{R,\rho} + \frac{1}{R} + 2\gamma < \sigma^*$. The contradiction shows that $W \subseteq C_0$, which concludes the proof of the Lemma. □

Appendix C. Proof of Theorem 3

For each $\eta > 0$, define $P_\eta = P(x : |x - x^*| \leq \eta)$ as the probability that the threshold realization is within η of x^*. If P does not have an atom at x^*, then, we can choose η_δ such that $P_{\eta_\delta} \leq \frac{1}{30}\delta$. Assume w.l.o.g. that $\eta_\delta \leq \delta$. Let

$$T_\delta = \left\{\tau : \frac{1}{N}\left|\{\tau_i : |\tau_i - x^*| \leq \eta_\delta\}\right| \leq \frac{1}{3}\delta\right\}.$$

The Law of Large Numbers implies that for sufficiently high N, $\text{Prob}(T_\delta) \geq 1 - \delta$.

Fix threshold profile $\tau \in T_\delta$. Let $I_0 = \{i : |\tau_i - x^*| \leq \eta_\delta\}$. Suppose that a is $\frac{1}{3}\eta_\delta$-fuzzy convention x^*. Let $I(g) = \left\{i : |\beta_i^a - x^*| > \frac{1}{3}\eta_\delta\right\}$ be the set of agents that is an equilibrium in game $G(g,\tau)$. Let $I = I_0 \cap I(g)$. Then, $\frac{1}{N}|I| \leq \frac{2}{3}\delta$. For each $i \notin I$, either

- $\tau_i > x^* + \eta_\delta$ and $\beta_i^a \leq x^* + \frac{1}{3}\eta_\delta$, which implies $a_i = a_i^* = 0$, or
  - $\tau_i < x^* - \eta_\delta$ and $\beta_i^a \geq x^* - \frac{1}{3}\eta_\delta$, which implies $a_i = a_i^* = 1$.

Hence, for any $i \notin I$, $a_i = a_i^*$. This concludes the proof of the Theorem.

References

Blume, L. E. (1993). The statistical mechanics of strategic interaction. *Games and economic behavior*, **5** (3), 387–424, publisher: Elsevier.

— (1995a). The Statistical-Mechanics of Best-Response Strategy Revision. *Games and Economic Behavior*, **11** (2), 111–145.

— (1995b). The statistical mechanics of best-response strategy revision. *Games and economic behavior*, **11** (2), 111–145, publisher: Elsevier.

— (2018). Population games. In *The economy as an evolving complex system II*, CRC Press, pp. 425–460.

Bollobás, B., Riordan, O. and Riordan, O. (2006). *Percolation*. Cambridge University Press.

Ellison, G. (1993). Learning, local interaction, and coordination. *Econometrica: Journal of the Econometric Society*, pp. 1047–1071, publisher: JSTOR.

— (2000). Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. *The Review of Economic Studies*, **67** (1), 17–45, publisher: Wiley-Blackwell.

Erdős, P. and Rényi, A. (1959). On Random Graphs. I. *Publicationes Mathematicae*, **6**, 290–297.

Friedman, D. (1991). Evolutionary games in economics. *Econometrica: Journal of the Econometric Society*, pp. 637–666.

Galeotti, A., Goyal, S., Jackson, M. O., Vega-Redondo, F. and Yariv, L. (2010). Network games. *The review of economic studies*, **77** (1), 218–244, publisher: Wiley-Blackwell.

Granovetter, M. (1978). Threshold models of collective behavior. *American journal of sociology*, **83** (6), 1420–1443.

Harsanyi, J. C. and Selten, R. (1988). A general theory of equilibrium selection in games. *MIT Press Books*, **1**, publisher: The MIT Press.

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

— and Yariv, L. (2007). Diffusion of behavior and equilibrium properties in network games. *American Economic Review*, **97** (2), 92–98.

Kajii, A. and Morris, S. (1997). The Robustness of Equilibria to Incomplete Information. *Econometrica*, **65** (6), 1283–1309.

Kandori, M., Mailath, G. J. and Rob, R. (1993). Learning, mutation, and long run equilibria in games. *Econometrica: Journal of the Econometric Society*, pp. 29–56, publisher: JSTOR.

Lee, I. H. and Valentinyi, A. (2000). Noisy contagion without mutation. *The Review of Economic Studies*, **67** (1), 47–56, publisher: Wiley-Blackwell.

Leister, C. M., Zenou, Y. and Zhou, J. (2022). Social connectedness and local contagion. *The Review of Economic Studies*, **89** (1), 372–410, publisher: Oxford University Press.

López-Pintado, D. (2008). Diffusion in complex social networks. *Games and Economic Behavior*, **62** (2), 573–590, publisher: Elsevier.

Monderer, D. and Shapley, L. S. (1996). Potential Games. *Games and Economic Behavior*, **14** (1), 124–143.

Morris, S. (2000). Contagion. *The Review of Economic Studies*, **67** (1), 57–78, publisher: Wiley-Blackwell.

— and Shin, H. S. (2006). Heterogeneity and Uniqueness in Interaction. *The Economy As an Evolving Complex System, III: Current Perspectives and Future Directions*, **3**, 207, publisher: Oxford University Press.

Mossel, E., Sly, A. and Tamuz, O. (2015). Strategic learning and the topology of social networks. *Econometrica*, **83** (5), 1755–1794.

Neary, P. R. (2012). Competing conventions. *Games and Economic Behavior*, **76** (1), 301–328.

— and Newton, J. (2017). Heterogeneity in preferences and behavior in threshold models. *Available at SSRN 3035289*.

Newton, J. (2021). Conventions under heterogeneous behavioural rules. *The Review of Economic Studies*, **88** (4), 2094–2118.

Peski, M. (2010). Generalized risk-dominance and asymmetric dynamics. *Journal of Economic Theory*, **145** (1), 216–248, publisher: Elsevier.

Ui, T. (2001). Robust equilibria of potential games. *Econometrica*, **69** (5), 1373–1380.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press.

WATTS, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, **99** (9), 5766–5771, publisher: National Acad Sciences.

YOUNG, H. P. (1993). The evolution of conventions. *Econometrica: Journal of the Econometric Society*, pp. 57–84, publisher: JSTOR.