Auton Agent Multi-Agent Syst (2014) 28:779–804
DOI 10.1007/s10458-013-9237-x

Emergence of conventions through social learning

Heterogeneous learners in complex networks

Stéphane Airiau · Sandip Sen · Daniel Villatoro

Published online: 20 October 2013

Abstract Societal norms or conventions help identify one of many appropriate behaviors during an interaction between agents. The offline study of norms is an active research area where one can reason about normative systems and include research on designing and enforcing appropriate norms at specification time. In our work, we consider the problem of the emergence of conventions in a society through distributed adaptation by agents from their online experiences at run time. The agents are connected to each other within a fixed network topology and interact over time only with their neighbours in the network. Agents recognize a social situation involving two agents that must choose one available action from multiple ones. No default behavior is specified. We study the emergence of system-wide conventions via the process of social learning where an agent learns to choose one of several available behaviors by interacting repeatedly with randomly chosen neighbors without considering the identity of the interacting agent in any particular interaction. While multiagent learning literature has primarily focused on developing learning mechanisms that produce desired behavior when two agents repeatedly interact with each other, relatively little work exists in understanding and characterizing the dynamics and emergence of conventions through social learning. We experimentally show that social learning always produces conventions for random, fully connected and ring networks and study the effect of population size, number of behavior options, different learning algorithms for behavior adoption, and influence of fixed

S. Airiau (✉)
LAMSADE, Université Paris Dauphine, Place du Maréchal de Lattre de Tassigny,
75775 Paris Cedex 16, France
e-mail: stephane.airiau@dauphine.fr

S. Sen
Department of Mathematical & Computer Sciences, The University of Tulsa,
800 South Tucker Drive, Tulsa, OK 74104, USA
e-mail: sandip@utulsa.edu

D. Villatoro
IIIA - Artificial Intelligence Research Institute, CSIC - Spanish Scientific Research Council,
Campus Universitat Autonoma de Barcelona, Bellaterra, 08193 Catalonia, Spain
e-mail: dvillatoro@iiia.csic.es

agents on the speed of convention emergence. We also observe and explain the formation of stable, distinct subconventions and hence the lack of emergence of a global convention when agents are connected in a scale-free network.

Keywords Norms · Conventions · Emergence

1 Introduction

Norms routinely guide the choice of behaviors in human societies. Conformity to norms reduces social frictions, relieves cognitive load on humans, and facilitates coordination. "Everyone conforms, everyone expects others to conform, and everyone has good reason to conform because conforming is in each person's best interest when everyone else plans to conform" [22]. Behaviors suggested by norms can therefore be considered to be in equilibrium as there is no incentive for any one individual to defy a norm if everyone else adopts it.

Norms in human societies range from fashions to tipping, driving etiquette to interaction protocols. Norms are ingrained in our social milieu and play a pivotal role in all kinds of business, political, social, and personal choices and interactions. They are self-enforcing: "A norm exists in a given social setting to the extent that individuals usually act in a certain way and are often punished when seen not to be acting in this way" [3]. Norms can therefore be substituted as external correlating signals or influences to promote coordination.

Boella et al. define a *normative multiagent system* as "a multiagent system organized by means of mechanisms to represent, communicate, distribute, detect, create, modify, and enforce norms, and mechanisms to deliberate about norms and detect norm violation and fulfilment". They recognize the importance of an *interactionist view on normative multiagent system*, i.e. a bottom up view in which a norm emerges without any enforcement, e.g., when the agents realize it is convenient for them to behave in a certain way. When a norm becomes established, i.e., when the norm is made explicit, there is a need for deliberating about it. When the norm has a positive outcome for the society, the norm can be adopted by establishing it as a rule or a law. In this case, one may reason about ways for enforcing the norm, i.e., by providing mechanisms to detect deviations from the norm and by determining possible actions/costs to punish deviators.

Not all the social norms, however, deal with the same kind of interaction scenarios. We observe that social norms like greeting (shaking hands, kissing, leaning towards each other, or a simple "hi!") pertain to different situations compared to, for example, the social norm of recycling. We also observe that social norms, though referring to the same concept, are defined using different terms in the literature, e.g. norms, social laws, conventions, social norms. In the multi-agent systems literature, several terms have been used for the same concept (convention, social norm, social law). However, those terms have been used as synonyms, often without clearly defining, delineating, or distinguishing between them. Considering norms as regularities in the behavior of agents, Coleman [10] defines two main types of norms: conventions and essential norms. These terms have been clearly differentiated in [40]. While *Conventional norms* (i.e. conventions) solve coordination problems where there exist no conflicts between the individual and the collective interests, as what is desired is that everyone behaves in the same way, *Essential norms* solve or ease collective action problems where there is a conflict between the individual and the collective interests.

In this paper, we study the emergence of *conventions* in a society of artificial agents through the repeated interactions between its members. Here, we use the term convention as we do

not consider any deontic character: we want to study the initial emergence of a collective behaviour that could be later on recognized as an explicit convention and could take deontic character. We believe that it is not advisable to decide on such conventions *offline* [33,42] (or at specification time). This is because not all the characteristics of the environment are perfectly known at design time and choosing a particular convention without such critically relevant specific knowledge could produce ineffective system performance. An online (or run-time) convention adoption process could provide a better-adapted convention, given ground realities, as the agents can tailor their decision to the current environment. In addition, the goals of the agents or the characteristics of the environment may change over time. In such situations, an offline design would require re-programming the agents, or at least put in place a mechanism to deliberate whether a change of conventions is required and how to implement it [34]. This would be not only computationally problematic in complex environment, but would also be costly and inefficient. Hence, it is important to study mechanisms that lead to emergence of a convention from online interaction experience.

While these considerations have merited in-depth study of the evolution and economics of conventions in social situations [14,33,45], our work is motivated by the following characterization by Young [46]: "we may define a convention as an equilibrium that everyone expects in interactions that have more than one equilibrium". This observation has particular significance for the study of conventions in the context of computational agents. Computational agents often have to coordinate their actions and such interactions can be formulated as stage games with simultaneous moves made by the players [17]. Such stage games often have multiple equilibria [25], which makes is a major impediment for achieving coordination. While *focal points*[1] [30] can be used to disambiguate such choices, they may not be available in all situations. Conventions can also be thought of as focal points evolved through learning [46]. Hence, the emergence of conventions via learning from experience in agent societies promises to be a productive research area that can improve coordination in and hence functioning of agent societies.

While researchers have studied the emergence of conventions in agent populations, they typically assume access to significant amount of global knowledge [14,19,45,46]. For example, all of these models assume that individual agents can observe sizable fraction of interactions between other agents in the environment. In that case, a convention may emerge as agents are mimicking the behavior of other agents. Shoham and Tennenholtz [33], however, provided results when the interactions are private. In particular, they prove that for a special class of games, a convention will emerge when the agents use a simple learning rule based solely on their private interactions. Delgado [12] also provided simulation results when the agents are using this learning rule and when the interactions between them are constrained. He considers an interaction graph where a vertex corresponds to an agent and an edge (i, j) represents the possibility of interaction between agent i and agent j. The main result of this work is simulation results that show that a convention always emerges when using a specific learning algorithm and that the topology has an effect on the speed of emergence of a convention.

Many situations involving many agents can be modelled as games where learning algorithms can then be used as decision making mechanisms. As a result, learning to play repeated games has been an active area of research in multiagent systems. Many learning algorithms have been designed and compared [1,27,41]. In particular, no single algorithm has been found to always outperform all the other algorithms. As a consequence, it is not clear that

[1] In game theory, a focal point is an equilibrium more likely to be chosen by the players because it seems special, natural or relevant to them, although other equilibria are equally good.

all agents would adopt the same learning algorithm in a decentralized environment. To study convention emergence in a multiagent system, it is then important to consider agents that use different learning algorithm and that learn from interaction with many different agent types. We study such scenarios in this paper.

To study the important phenomenon of emergence of conventions via private interactions, we use the following interaction framework. We consider a population of agents, where, in each interaction, each agent is paired with another agent randomly selected from the set of agents it is connected with in a network. Each agent learns from its interactions concurrently and over repeated interactions with randomly selected neighbors. We refer to this kind of learning *social learning* to distinguish from learning in iterated games against the same opponent [15].

In previous work on learning in repeated games, the opponent is fixed but in our work, the opponent varies from iteration to iteration. In addition, it is unlikely that all agents, in a decentralized environment, will all use the same learning algorithm. Even if they did, they may not use the same set of parameters. It is unclear, a priori, if and how a social convention will emerge from such a social learning framework. We believe this is a richer and more realistic scenario for interaction, behavior learning, and convention emergence in artificial agent societies.

1.1 Our contribution

Our focus is on the early adoption of a convention, which, later on, could receive some deontic character (what one should do). We are also only concerned with a society of artificial agents, so we do assume some level of rationality and we do not try to model emergence of conventions in human societies. This implicitly assumes some common preferences (e.g. all agents prefer to spend less time than more time at an intersection), but we do not believe this as a particularly restrictive assumption. If a system designer has not recognized that a convention could benefit the agents, we want to know (1) whether agents could adapt their behavior to take advantage of this opportunity and (2) whether the collective behavior is stable. Our assumption is that agents are able to recognize a social situation and they learn to adopt a convention that will facilitate coordination or problem solving for that social situation.

Our first contribution is to extend the results of Shoham and Tennenholtz [33] and Delgado [12] in our richer and more realistic model of agents society. Another contribution is the investigation of conditions in which different conventions may emerge in different parts of the population. This is possible when two or more isolated population groups interact infrequently: with significantly more interactions within members of the same group, infrequent interaction with outsiders is unlikely to have a decisive effect on the convention adopted by members of a group. Consequently, it is possible that the different groups develop different conventions. Delgado [12] and Shoham and Tennenholtz [33] did not observe such phenomenon. We observed this fact when studying two populations that interacts unfrequently as well as studying emergence of conventions in one specific agent interconnection topology (scale free networks).

Our experimental results and concomitant analysis throws light on the dynamics of the emergence of conventions via social learning with private interactions. We also investigate a number of key related issues: the effect of population size, number of available actions, multiple populations with limited inter-population interactions, heterogeneous population with multiple learning algorithms, effect of non-learners in shaping convention adoption, conventions for social dilemmas, etc.

The rest of the paper is organized as follows: in Sect. 2 we review the various related works and motivate the research presented in this paper; in Sect. 3, we present the simulation model used for the experiments; experimental results showing that a convention emerges are presented in Sect. 4, whereas we analyze, in Sect. 5, the special case where multiple conventions, or subconventions, may emerge in different parts of the population; finally conclusions and future work are presented in Sect. 6.

2 Related work

The need for effective norms to control agent behaviors is well-recognized in multiagent societies [7,37]. In particular, norms are key to the efficient functioning of electronic institutions [16]. Most of the work in multiagent systems on norms, however, has centered on logic or rule-based specification and enforcement of norms [13,37]. Similar to these research, the work on normative, game-theoretic approach to norm derivation and enforcement also assumes centralized authority and knowledge, as well as system level goals [6,7]. While norms can be established by centralized dictat, a number of real-life norms evolve in a bottom-up manner, via "the gradual accretion of precedent" [46]. One potential application where this research can apply is in the decentralized emergence of "language" (or shared ontology), as it was initially sketched in [21], where authors study how a population of random agents can emerge with a shared vocabulary wihtout central entities ruling this process.

Some models of emergence of conventions have appeared in the game theory literature. One common feature is that a social situation is represented by a game. Agents have some preferences between the various outcomes of a given social situation (we will assume that agents either strictly prefer one outcome over another or they are indifferent between two outcomes). In our work, since we do not consider deontic character, we chose to use the term conventions, while many authors use the term norm and in the following, we respected their choice.

To facilitate the modelling process, one can use interval scales to represent the preferences. When the situation involves two agents and the agents have the choice between p discrete actions, the game can be represented in a table form such as the one presented in Table 1 for the case of $p = 2$ actions. For a game represented by a table R, the rows and the columns represent the action choices for the "row" and the "column" player respectively. The entry $R(a, b) = (u, v) \in \mathbb{R}^2$ describes the outcome when the "row" player picks action a and the column player picks action b: the row player obtains a utility of u and the column player a utility of v. The example of Table 1 models many situations where the agents must coordinate to receive higher payoffs [33]. If the agents do not take the same action, at least one of the agent experience a negative utility. When they choose the same action, at least one outcome brings positive utility to both agents. This generic type of game can model coordination (e.g. $x = y = 1$ and $u = v = -1$) as well as more complex scenarios such as the well known Prisoners' Dilemma game (e.g. $x = 1, y = -2, u = 3, v = -3$).

More complex representations of a situation are possible. For example, Verhagen [38] used a decision tree to model situations, which allows the computation expected utilities for

Table 1 Social agreement game [33]

$\begin{pmatrix} x, x & u, v \\ v, u & y, y \end{pmatrix}$	$x, y, u, v \neq 0$ either $x > 0$ or $y > 0$ either $u < 0$ or $v < 0$ if $x > 0$ and $y > 0$, then $x = y$.

the different outcomes. Some learning mechanism will make similar computations as they are based on maximizing expected utility.

One common assumption is that agent can base their strategy upon the observation of other agents. Young [45], for example, is not interested in learning by individual agents, but on agents observing the population and taking decisions based on these observations. In his model, each agent observes the last m interactions and bases its decision by building a Markov chain from k random samples selected from these m past interactions. These k samples correspond to an agent asking/observing k people. Once the agent has taken a decision, it can die or leave the environment, ensuring that convention emergence is not due to learning or reputation effect, i.e., the convention emerges by mimicry. Epstein considers a model where agents are organized in a ring and can observe the behavior of neighbors [14]. The strategy adopted by an agent is the strategy used by the majority of its neighbors. To gear towards a stronger consensus, the agent will compare the majority decisions of two sets of neighbors: the first include neighbors situated within a distance of r and the second add neigbors that live within a distance of $r + 1$. When the decisions are the same, the agent will follow it, otherwise, the agent will increase the radius r.

Axelrod [2] used an evolutionary approach: agents "observe each other, and those with poor performance tend to imitate the strategies of those they see doing better". Axelrod thus studies a behavioral approach which requires observation of other agents. One interesting feature of this work is that to ensure stability, agents can punish agents that do not follow the norm. We do not need such punishment as the use of a learning mechanism leads to a rational behavior: unless an agent is exploring, her decision will maximize expected utility.

Walker and Wooldridge performed simulations of norm emergence in two-player coordination games [42]. No payoff is defined, but an interaction is a 'success' when both agents choose the same strategy and is a 'failure' otherwise. The agents record each interaction, i.e., a pair containing the strategy of both players. Walker and Wooldridge introduce different variants of the following learning rule called *simple majority* rule. An agent counts the number of times a given strategy was used by an opponent, and it picks the action with the maximum count. In the first variant, the memory of an agent is erased when it changes strategy. In the next two variants, agents can communicate their memory: in the first case, agents are divided into two groups and they exchange their memory when they interact with a fellow group member, in the second case, an agent shares interactions only when it has been 'successful' using a particular strategy, and will only share the interactions where the use of that strategy were successful. They also consider different quotas for the majority rule, i.e., in a double majority, an agent will change from strategy σ to strategy σ' only when the number of observation of σ' is twice the number of observation of σ.

In all these models, an agent may observe the interactions between other agents. In some other models, agents may even communicate their model of the situation. This is for example the case in [38] where agents have a model for their own model and learn a model of the group behavior.

In the model we use in this paper, agents rely exclusively on their own personal experience to choose a behavior. This is similar to the model of Shoham and Tennenholtz [33]. A norm is modeled as as an equilibrium of a stochastic game, not necessarily a Nash one (e.g. using a norm could dictate playing cooperate in the Prisoners' Dilemma, producing an outcome that is Pareto optimal even if it is not a Nash equilibrium). They study *n–k–g stochastic social games* where n is the set of agents and g is a k-person game. At each time step, k agents are selected from a uniform distribution over the n agents. Agents learn to play the game without knowing the identity of the set of agents that also play the game (*obliviousness* property) and can use their private history (*local* property) or histories of other agents (*semi-local* property).

This corresponds to our model in the case where all agents are equally likely to interact with one another. Shoham and Tennenholtz introduce the *highest cumulative reward (HCR)* rule, a simple local learning algorithm where an agent plays the action that has the largest total payoff in its memory over the last m iterations. The paper focuses on n–2–g stochastic games, which corresponds to the case we are studying. For the class of n–2–g games where g is a social agreement game as presented in Table 1, they prove that *a norm is guaranteed to emerge if all agents use HCR*. Their other contribution is a simulation of the convergence speed of HCR, studying the effect of variants of the algorithm, e.g., changing the frequency of strategy update, resetting memory whenever an agent changes strategy, using different memory sizes.

Though we do not provide convergence guarantees in our work, we do provide extensive simulations showing that a convention emerges most of the time even when individual agents use different types of learning algorithm. This and other studies described above have considered that agents interact with randomly selected partners from the entire population. We also show that a unique convention may not emerge in the population and different conventions may emerge in different groups in the population when the agent inter-connections form some specific topologies.

Some other researchers have also considered that the interactions are constrained by an underlying interconnection topology [12,20]. The learning rules mentioned above can still be applied in such situations. For example, the HCR rule is also investigated by Kittock [20] in the context of simple network topologies whereas Delgado [12] considers both HCR and majority rule in the context of complex networks, including scale-free graphs. Recent results show how the topology of interaction plays a fundamental role in the emergence of convention-like norms, which authors refer to as *the spread of innovation* [23]. In practice, computational agents are connected through specific networks with known topologies, and hence it is necessary that we test the emergence of norms under common agent interconnection topologies.

Different performance criteria have been used to measure the rate, success, and cost of norm formation from experience. Walker and Wooldridge [42] consider the number of changes of strategy by one agent: as changing strategy is considered costly, the less number of changes, the better. Kittock [20], Delgado [12], and Shoham and Tennenholtz [33] evaluate the speed of emergence measured in terms of the number of time steps before a norm is established. Both Kittock and Delgado recognizes the emergence of a norm when 90 % of the agents use the same strategy. Shoham and Tennenholtz recognizes emergence only when a norm is unanimously adopted, i.e., all agents use the same strategy.

Trying to provide agents with other mechanisms, Urbano et al. [36] investigated how a new strategy update rule (*Recruitment based on Force with Reinforcement*), which promotes a dynamic creation of a hierarchy, speeds up the convention emergence. Their proposed strategy update rule functions by providing agents with a measure of *force* that increases with every successful interaction, and in case of unsuccessful interactions, the agent with the lower force copies the strategy and force of the winner. We find this function similar to the SLACER algorithm [18]. Authors prove the efficiency of their new strategy update rule on different topologies: regular graphs, an unconventional implementation of a random graph with uniform degree distribution, scale-free (developed following the Albert–Barabasi model), and Small World (constructed using the Watts–Strogatz model) networks. However, convergence rate is still fixed to 90 %. Their results are in concordance with those obtained by Kittock in [20] where he proved the efficiency of trees structures, that are those dynamically created with the concept of force.

We use agents that use multiagent reinforcement learning algorithms [28,35] to adapt their behavior based on their interactions with other agents in the society. Most multiagent reinforcement learning literature involve the same two agents repeatedly playing a stage

game and the goal is to learn policies to reach preferred equilibrium [29]. Another line of research considers a large population of n agents learning to play a cooperative n-player game where the reward of each individual agent depends on the joint action of all the agents in the population [44]. The goal of the learning agent is to maximize an objective function for the entire population, the world utility. The social learning framework we use to study convention emergence in a population is distinct from both of these lines of research. We are considering a potentially large population of learning agents. At each time step, however, each agent interacts with a single agent, chosen at random from its neighbors in the interconnection graph. The payoff received by an agent for a time step only depends on this interaction, and is independent of the behavior of any other agent. In addition, the opponent, whose identity is not observed, is likely to be different in different interactions. It is not clear *a priori* if the learners will converge to stable, useful policies in this situation, but our results confirm that conventions indeed emerge through such social learning!

3 Social learning framework

We assume that agents have higher level goals. When they try to achieve these goals, they may have to interact with the other agents present in the environment (maybe robots are about to meet on the road or arrive together at the same door). Conventions improve efficiency as agents would not require explicit communication to deal with these specific situations. In the following, the conventions are implicit: agents adopt a personal stable behaviour that can be seen as a convention at the macroscopic level. It is possible that agents could reason about their own behavior later on and discover that they are using a convention and that they could make it explicit. But in this paper, we only deal with the early choice of a specific behavior.

In the following, we present our framework to study the emergence of conventions in a society of interacting learning agents. We use a normal-form game to represent a social situation between two agents. We also assume that agents are located in a fixed interaction topology that constrains their interactions. We use an interaction protocol in which, at each iteration, an agent meets another randomly chosen agent from its neighbors in the interconnection graph. We assume that interacting agents do not learn separate behaviors for interacting with different agents in the population. For a large population, it would take a long time to learn and it would not be efficient in an open environments where agents can enter and leave the system over time. We assume that agents are able to recognize a type of social interaction and are able to access past experience with other agents. Using this data, they use one of the several learning algorithms to adapt their behavior based on interaction experience.

It is out of the scope of this paper to discuss the semantics associated to the definition of a conventions. For example, if a behaviour is explicitly defined as a convention, this does have an effect on the compliance by humans. In this work we focus on the initial choice of a behaviour and we analyze the effect of the social learning and the topological distribution of agents on the emergence of conventions.

3.1 Modeling a social interaction

We consider a population N of interacting agents. In theory, a convention could involve an interaction between many agents at the same time, but in practice, many conventions prescribe the expected behavior between two agents, e.g., which hand to extend in greeting or which side of the road to drive on. Consequently, we restrict our study to the case of *bilateral* interactions. One specific example of social situation for convention emergence that we use

as a running example is that of learning "rules of the road". For example, we will consider two representative convention emergence scenarios: (a) which side of the road to drive on, and (b) who yields if two drivers simultaneously arrive at an intersection from neighboring roads[2].

The first assumption is that an agent is able to abstract a situation and recognize a specific type of interaction. Two agents may have symmetrical view of the interaction (e.g., in the case of choosing which side of the road to drive on) or different one (e.g., in the intersection problem, one driver sees a car on its left and the other sees a car on its right). For our definition of a social interaction, we focus on one viewpoint and each agent plays a specific role: first role (or row role) and second (or column) role. Of course, we consider that each agent can experience both roles. The agent has to select an action in the set $\mathscr{A}_r$ when she has the row role (respectively in the set $\mathscr{A}_c$ for the column role). We use a normal-form game to model the preferences between the different outcomes of the interaction. Each agent models the situation with a matrix G_i of size $|\mathscr{A}_r| \times \mathscr{A}_c|$. We now formally define a social interaction and its outcome.

Definition 1 (*Social interaction*) A social (bilateral) interaction is a tuple $\langle N, \mathscr{A}_r, \mathscr{A}_c, (G_i)_{i\in N}\rangle$ where $\mathscr{A}_r$ (resp $\mathscr{A}_c$) is the set of actions available to the row role (resp. the column role) and G_i is the payoff matrix of agent i such that:

- agent i gets $G_i(a_r, a_c)$ when i is the row agent and chooses action a_r and the other agent is the column agent that chooses a_c.
- agent i gets $G_i(a_r, a_c)$ when i is the column agent and chooses action a_c and the other agent is the row agent that chooses a_r.

We assume that all agents have a similar understanding of the social situation: all the agents share the same ordinal preference over the different outcomes of the game, but they may have different cardinal preferences. In the intersection example, this means that all agents prefer to be the one that does not stop. This assumption excludes from consideration situations where agents that have different preferences between the outcomes (e.g. a risk averse agent may prefer the outcome where she yields and the other agent goes). Since our goal is to model convention emergence in artificial agents societies, we believe this assumption is not a severe restriction. Note however that since we use cardinal utility, we allow the actual payoffs to differ from agent to agent, and we allow indifference: if some agents strictly prefer outcome a over outcome b, then no agent strictly prefer b over a, but some may be indifferent over them. In most experiments though, we have used the same game for all the agents.

A possible pair of payoff matrices for the intersection problem is presented in Table 2. While each player has the incentive of not yielding, myopic decisions by both can lead to undesirable accidents. Both drivers yielding, however, also creates inefficiency. When one player "go" and the other "yields", the player that yields gets a lesser payoff since it is losing some time compared to the other player. The players know whether they are playing as a row or a column player: the row player sees a car on its right, and the column player sees a car on its left. The action choices for the row player are to go (G) or yield to the car on the right (Y_R), and they are go (G) or yield to the car on the left (Y_L) for the column player. In the table, we present the matrix for two agents i and j. The utility for agent j after having an accident is much lower than for agent i (maybe it values its car much more). They have the same utility for going when the other yield, but agent i values more, due to different intrinsic biases, the act of yielding when the other goes. Agents can also sample their cardinal representation

[2] It might seem that "rules of the road" are always fixed by authority, but historical records show that "Society often converges on a convention first by an informal process of accretion; later it is codified into law" [46].

Table 2 Stage games for two agents i and j corresponding to an intersection problem

	G	Y_L
agent i		
G	-1	3
Y_R	2	1
agent j		
G	-1	3
Y_R	1	1
G	-1	γ
Y_R	β	α
distribution of games		
$\alpha = rand(\{0, 1\}) + rand([0, 1])$		
$\beta = \alpha + 2 * rand([0, 1])$		
$\gamma = \beta + rand([0, 1])$		
$rand([0, 1])$ draws a real from the uniform distribution on the interval [0, 1]		
$rand(\{0, 1\})$ draws an integer from the uniform distribution on the set{0, 1}		

Table 3 Stage games for two agents i and j corresponding to choosing which side of the road to drive on, a coordination game

	L	R
agent i		
L	4	-1
R	-1	4
agent j		
L	2	-10
R	-10	2

from the distribution of games described in the matrix on the right of Table 2. Of course, many other distributions satisfy the ordinal preferences. Note that the expected payoff of an agent that draws her preferences from this matrix is the one of agent i.

The payoff matrices for the problem of choosing which side of the road to drive on is presented in Table 3. It is an example of coordination: if both agents choose the same side (from their point of view), they safely continue on their route; otherwise, they risk a frontal crash. In the shown payoff matrices, agent j fears a crash more than agent i, again due to differing intrinsic bias, with a much higher penalty.

Finally, in this paper, we will assume that an interaction is private. In the example of the intersection, an agent that is near the intersection could potentially observe how the agents interact, but we will assume this is not the case. Our goal in this paper is to show that only the information from personal interactions is sufficient for a convention to emerge.

3.2 Interconnection topology

Agents are connected by a interconnection graph $\mathcal{G}$ of a fixed topology that restricts their interactions only with their direct neighbors. Such physical or spatial interaction constraints

or biases have been well-recognized in the social sciences [26] and, more recently, in the multiagent systems literature [31]. More pertinently, such interaction constraints have also been used in the context of convention emergence [12]. We use the following agent interconnection topologies in our experiments:

Fully connected networks: In this topology, each agent is directly connected with all other agents, i.e., there is no constraints on the interaction between any two agents.

One-dimensional lattice with neighborhood size k: This topology provides a structure in which agents are connected with their k nearest neighbors. Different values of the neighborhood size, k, produces different network structures; for example, when $k = 2$ the network will have a ring structure, as in Fig. 1b, and agents will only be connected with their direct neighbors on the left and right. On the other hand, when $k = |N|$, the network is a fully connected network, as in Fig 1a, where each agent is connected with all other agents.

Scale-free networks: In this topology, the number of connections k originating from a given node exhibits a power law distribution $P(k) \approx k^{-\gamma}$ where γ is typically in the range $2 < \gamma < 3$. As a result, few agents have many connections, and many agents have few connection. The internet or the citation networks are commonly cited example networks that have properties of scale-free networks. The diameter (average minimum distance between pairs of nodes) of such networks is significantly smaller than the *one-dimensional lattice*. We generated scale-free networks using the algorithm proposed by Barabasi and Albert [5], with $m_o = 2$, $m = 1$, $p = 1$ and $q = 0$. These parameters tune the generation of the scale free network, starting with 2 nodes (because the value of m_o), adding one node (because the value of m) each iteracion (because the value of p) and not rewiring any link (because the value of q).

For the intersection problem, two agents that live relatively close to each other should be connected, as they are likely to meet each other at an intersection. Somebody that lives far away is much less likely to encounter the first two agents.

3.3 Interaction protocol

So far, we have a population N of $n = |N|$ agents located in a graph $\mathscr{G}$ that faces a social situation involving two roles, with their corresponding action set $\mathscr{A}_r$, $\mathscr{A}_c$, and each agent i models the social situation with a game G_i. Hence, we represent the social situation as a tuple $\langle N, \mathscr{G}, \mathscr{A}_r, \mathscr{A}_c, G_1, \ldots, G_n \rangle$.

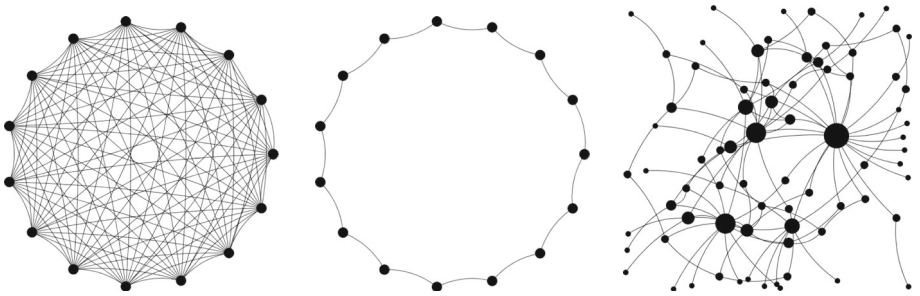

Fig. 1 Underlying topologies. **a** Fully connected network. **b** Ring network or one-dimensional lattice with neighborhood size 2. **c** Scale-free network

The simulation of the system progresses in discrete, synchronous steps. At each iteration, many distinct pairs of agent are randomly generated. To make a pair, an agent is first chosen from the set of agents that has not already been selected, and is paired with a randomly selected neighbor that has not yet been selected. This selection process is iterated until no more pairs can be formed, i.e., when there are no more neighbor agents that have not been selected. Algorithm 1 provides a precise description of this protocol.

Input: N: set of agents
$\mathscr{G} \subseteq N \times N$: symmetric relation modeling neighbours
for *a fixed number of epoch* **do**
  $A \leftarrow N$ //Initialize the available agents with the entire population
  repeat
    randomly select an agent from the available agents: $i \in A$;
    randomly select a neighbor of i that is available: $j \in A \cup \{j \mid (i, j) \in \mathscr{G}\}$;
    remove i and j from the set of available agents: $A = A \setminus \{i, j\}$;
    With probability $\frac{1}{2}$ draw $(p_{row}, p_{col}) = (i, j)$, else draw $(p_{row}, p_{col}) = (j, i)$;
    let p_{row} to select an action r in $\mathscr{A}_r$;
    let p_{col} to select an action c in $\mathscr{A}_c$;
    send the joint action (r, c) to both p_{row} and p_{col} for policy update;
  until *no pair of agents is available:* $\nexists (i, j) \in A^2 \mid (i, j) \in \mathscr{G}$;

Algorithm 1: Interaction protocol.

In the case of a fully connected graph, this algorithm will produce $\frac{N}{2}$ pairs at each iteration, and all the agents will learn at the same speed. However, when agents do not have the same number of connections, the agents that have more neighbors are more likely to experience an interaction at each iteration. This selection imbalance introduces a bias: agents with more connections will accumulate more experience and will learn faster than agents with a small number of connections. Because of the topology of the graph, agents with many connections will also have a greater influence on others. In our setting, they are likely to learn faster and set the trends for neighboring agents that have few connections.

To recap, some agents may not have an interaction during an epoch in our protocol as all their neighbours are no longer available. Note that one could study some variants of the protocol for guaranteeing that all agents interact at least once at each epoch. For example, one could allow one agent to have multiple interactions during the same epoch, in which case one can make the choice to update the learning algorithm once or after every interaction. We will not study these variants in this paper.

3.4 Convention

The behavior of an agent in repeated play of a bilateral stage game is characterized by its actions when it plays respectively the row and the column role. More formally, given a social interaction $\langle N, \mathscr{G}, \mathscr{A}_r, \mathscr{A}_c, G_1, \ldots, G_n \rangle$, the behavior of an agent i is a pair (r^i, c^i) that consists of a pure strategy $r^i \in \mathscr{A}_r$ for the row role and a pure strategy $c^i \in \mathscr{A}_c$ for the column role (the strategies are the actions available to the agents). A convention corresponds to an equilibrium strategy profile for all pairs of agents in the population:

Definition 2 *(convention)* For a social situation $\langle N, \mathscr{G}, \mathscr{A}_r, \mathscr{A}_c, G_1, \ldots, G_n \rangle$, we say that the population uses a *convention* when for all pairs of agents (i, j), we have both that (r^i, c^j)

is an equilibrium for the game (G_i, G'_j) and (r^j, c^i) is an equilibrium for the game (G'_i, G_j) where G' is the transpose of G.

Not all equilibria are conventions, and in the following, we will consider that a convention is a pure strategy Nash equilibrium. The use of a Nash equilibrium ensures that the equilibrium is stable: knowing that other agents are following the convention, a given agent is incentivized to follow it as well. We consider only pure strategies as, in practice, conventions are pure strategies.

When a game possesses a unique Nash equilibrium in pure strategies, rational agents should play this strategy. But this is not an interesting situation as there is no other option attractive to rational agents. In the rest of the paper, we will consider that the games G_i used by the agents have multiple Nash equilibrium in pure strategies. In the example of the intersection problem from Table 2, ideally, we would like conventions like "yield to the driver on the x", x being either left or right, which serves all drivers in the long run. Hence, the dilemma is resolved if each member of the population learns to "yield" as a row (column) player and "go" as a column (row) player. Note that for a social convention to evolve, all agents in the population has to learn any one of the following policy pairs: (a) (row:G, col:Y_L), i.e., yield to the car on the left, or (b) (row:Y_R, col:G), i.e., yield to the car on the right.

3.5 Learning algorithms

We assume that the basic decision making mechanism of an agent is a learning mechanism. This allows the agents to adapt their behaviors to their current environment and reduces the need for the system designer to specify precise parameters for each environment. In this paper, we will further assume that agents try to learn a behavior at the level of a social interaction. For example, an agent does not try to learn a behaviour that depends on the specific agent she interacts with. If an agent fails to learn an appropriate behavior at this level (not all situations may be solved by a normative behaviour), we would assume that the agent would refine her model of the situation (e.g. consider different sub-situations or learn some exceptions). Then, the agent would be able to recognize a given situation and use the appropriate data to make a decision. Discussion of recognizing situations falls outside of the scope of this paper and we will concentrate on a single social situation.

With the social learning framework, there is no theoretical guarantee that a convention does emerge and stabilize. The goal of the agents is not to learn or discover a convention. Rather, their goal, as rational agents, is to choose decisions to maximize their expected utility. We wanted to test wether it was possible for learners to choose a correct behaviour, and we chose reinforcement learning as our tool for decision making. Over the interactions, the agents learn to expect others to behave in a certain way.

We are studying the emergence of a convention in a population of interconnected interacting agents. Each agent uses a learning algorithm to learn, from accumulated experience, how to behave in each role of the social situation. We will at first assume that the agents do not have any initial bias towards a particular equilibrium. We want to observe whether the population is able to learn the same behavior, i.e., whether, in the long run, the population adopts a convention.

To learn a useful behavior, an agent needs to first explore its options and subsequently exploit its accumulated knowledge. Even when the behavior of other agents appears predictable, an agent will need to explore periodically to ensure that it is not using a sub-optimal strategy. It is also important in case of a change in the environment or if they change environments. This is particularly important for open and dynamic agent societies that are of

interest to us and to a large percentage of researchers in the multiagent systems community. Consequently, when agents are learning, we cannot expect to observe a point in time subsequent to which all agents will always choose the same strategy. To identify the emergence of a convention, we will use the following definition:

Definition 3 *(Convention emergence)* For a social situation $\langle N, \mathscr{G}, \mathscr{A}_r, \mathscr{A}_c, G_1, \ldots, G_n \rangle$, A *convention has emerged* when the strategy profile $(r, c) \in \mathscr{A}_r \times \mathscr{A}_c$ is played by 95 % of the population in a given iteration.

This definition is similar to the ones from [33], and from [12,20] (though in the latest the convergence rate used in the experiments is 90 %). One could achieve better performances if agents slowly reduces, and ultimately stops, exploration. We chose not to do so to prove that the results do not occur because of a special exploitation strategy.

Note that each agent must learn how to behave for each role of the social situation. In this paper, we will assume that both roles are learnt *independently*. Of course, this issue is relevant for interactions that are not symmetrical (e.g. the intersection problem: one driver sees a car on its left, the other driver sees a car on its right). For symmetric problem, we would simply need a single learning algorithm, but, as we discuss now, we can also learn it with two learning algorithms.

Learning independent behaviour for each role has important implications when a population contains only two agents: combinations of different Nash equilibrium may also emerge. For example, for the example intersection problem and for a population of two agents, it can be the case that one learner learns to "go" both as row and column player and the other player learns to yield in both situations (if the agent that goes is an ambulance, this would be the right behavior to learn). Although not "fair", this situation is possible in our framework since each agent independently learns to play as a row and a column player. But one could not refer to this process as social learning when only two agents are present.

When a third agent is introduced, as the agents do not know the identity of the opponents, no agent can any longer benefit from always choosing "go". This is because all other agents must always "yield" to the "go" agent, and then those agents will receive relatively poor utility when playing each other. As a result, they will also learn to "go" sometimes. To optimize performance they will have to learn to settle to a convention which everyone else also follows.

We use three different learning algorithms that are well-known in the learning in games and multiagent learning literature:

Fictitious play (FP): FP is the basic learning approach widely studied in the game theory literature [15]. The player keeps a frequency count of its opponent's decisions from a history of past moves and assumes that the opponent is playing a mixed strategy represented by this frequency distribution. It then chooses a best response to that mixed strategy, with the goal of maximizing expected payoff. This player models its opponent's behavior and tries to respond optimally. FP learns to respond optimally to an opponent playing a stationary strategy. Note that in our framework, the opponent is always changing, hence the history of the agent using FP and its current opponent's behavior may differ. Convergence is not guaranteed in such an environment.

Q-learning [43] with ϵ-greedy exploration: Q-learning has been widely used in single and multiagent systems, but converges only to optimal pure strategies is guaranteed only in a single-agent setting. This algorithm has been developed to learn an optimal policy in a Markov decision process. We use the ϵ-greedy exploration scheme: with probability $1 - \epsilon$ the agent follows the recommendation of Q-learning, but with a probability ϵ, the agent takes an action at random. Again, there is no guarantee of convergence in our social learning environment.

Win or learn fast-policy hill climbing (WoLF-PHC) [9]: The idea behind WoLF is to quickly adapt when losing but be cautious when winning. WoLF extends policy hill climbing, which itself extends Q-learning by adding the ability to learn mixed strategies. Though WoLF is guaranteed to converge to a Nash equilibrium in repeated play of a 2-person, 2-actions game against a given opponent, it is not clear whether convergence is guaranteed in social learning.

Though there are other competitive multiagent reinforcement learning algorithms (e.g. GIGA-WoLF [8], M-Qubed [11], CJAL [4], etc.), we believe the set of algorithms we used produces representative results for convention emergence through social learning.

4 Experimental results: emergence of a convention

Previous research has shown that a convention emerges in a society of adaptive agents, but they assumed that all agents were using the same algorithm. We now provide new evidences of convention emergence using our more realistic framework of social learning. We study the emergence of conventions under different settings: in different games, with different population sizes and number of actions. We also study the effect of the learning algorithms used and the effect of fixed agents that cannot (or do not want to) learn. In the first set of results, we will assume that the agents are connected through a complete graph[3]. In the last subsection, we will assume that agents are connected using a one-dimensional lattice.

4.1 Example of a social dilemma

One typical example of the use of convention is to resolve social dilemmas. A straightforward example of this is when two drivers arrive at an intersection simultaneously from neighboring streets (See Sect. 3.1). For this experiments, all agents use the game of agent i in Table 1.

Our experimental results show that a uniform convention always emerges in a population of three or more agents. For example, in a population of 200 agents using WoLF with ϵ-greedy exploration: the agent chooses to follow the action recommended by WoLF with a probability of $1-\epsilon$, otherwise, it draws an action from a uniform distribution. We ran 1,000 runs, and we observed that the population converged to the "yield to the left" convention 506 times, and "yield to the right" convention 494 times. We present the averaged dynamics of the payoffs and the frequency of the joint action during learning in Fig. 2. From the dynamics we can see that at first the agents avoid the collision and prefer to yield. Then, one agent notice that it can exploit this situation by choosing to "go" as the other one is yielding. Depending on who notices this first, the population converges to one convention or the other. Note that the plot in Fig. 2 is averaged over all the runs, which explains why the (G, Y_L) and (Y_R, G) appear almost 50 % of the time. The presence of the other joint-actions is due to the ϵ-greedy exploration. At the end of the simulation, we observe that the two norms occurs 48.5 and 47.4 % of the time respectively and the other two joint action occur about 2 % of the time, which is in accordance with playing a fixed pure strategy $1-\epsilon = 0.96$ of the time and choosing randomly an action the rest of the time[4].

[3] These results were published in [32].

[4] If the norm (G,Y_L) has emerged and all agents play ϵ-greedy with $\epsilon = 0.04$, we will observe the outcome (G,Y_L) with a probability of $\left(.96+\frac{.04}{2}\right)^2$, (G,G) and (Y_R,Y_L) with a probability of $(.96+\frac{.04}{2})\cdot\frac{.02}{2}$ and (Y_R, G) with a probability of $\left(\frac{.02}{2}\right)^2$. Overall, we have a probability of 0.4804 to observe each (G,Y_L) and (Y_R,G) and a probability of 0.0196 to observe (G,G) and (Y_R,Y_L).

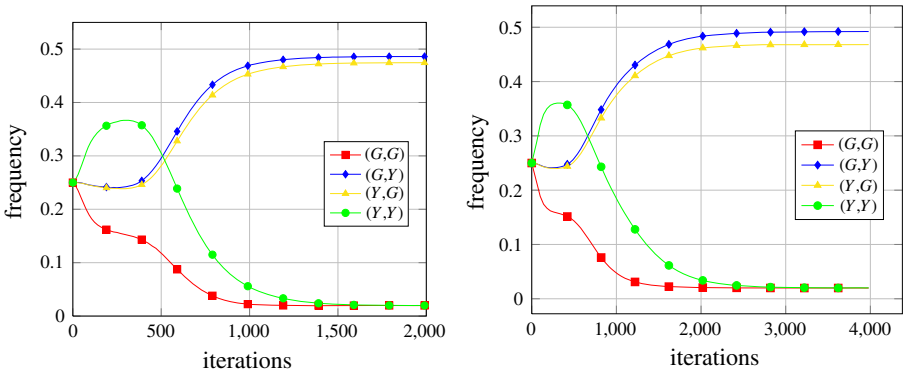

Fig. 2 Social dilemma game with 200 agents using WoLF, averaged over 1,000 runs. Agents use the game of Table 2 player i. (G, Y_L) emerges 534 times and (Y_R, G) 466 times. Agents sample their preferences from Table 2. (G, Y_L) emerges 506 times and (Y_R, G) 494 times

Using the same settings, we ran a simulation where all the agents have the same ordinal preferences, but may have different cardinal representation of these preferences. At the begining of each run, each agent draws her cardinal representation from the distribution of games presented in Table 2. So, all agents may experience different cardinal payoffs for each joint-action. On expectation, an agent has the same payoff as in the previous simulation (on average, the payoffs are the ones of the game of player i in Table 2 used in the previous simulation). We always observe the emergence of a norm. The dynamics is very similar to the previous simulation, except it is slower. For some agents, except for the value of the (G, G) state, the differences between the other payoffs may be small and agents may take some additional time to learn the small differences.

These results confirm that only private experience is sufficient for the emergence of a convention in a social learning framework. This is in contrast with prior work on convention evolution which requires agents to have knowledge about non-local interactions between other agents and their strategies [14,19,45].

4.2 Influence of population size, number of actions, and learning algorithm

The time required for the emergence of a convention in a society of interacting agents, measured by the number of interaction periods before most agents adopt the convention, depends on several factors. Here we study the influence of the size of the population, the learning algorithm used, and the number of actions available to the agents.

First we consider the effect of population size. With a larger population, the likelihood that two particular agents interact decreases. Hence the variety of opponents as well as the diversity of personal interaction history increases with the population size. As a result, the population takes more time to evolve a convention. In Fig. 3, we present the dynamics of the average agent reward for the social dilemma game in a population of agents using WoLF with different population sizes: with more agents, it takes longer for the entire population to converge on a particular convention. In the real world, it is well-known that tight-knit, small societies, groups, clans develop eclectic conventions that are often not found in larger, open societies.

Our hypothesis is that learning is facilitated when the agents have a very similar history. With a small population, the same agents will meet often and they will have a similar history. When the population becomes larger, it takes longer to meet the same agent on average. The

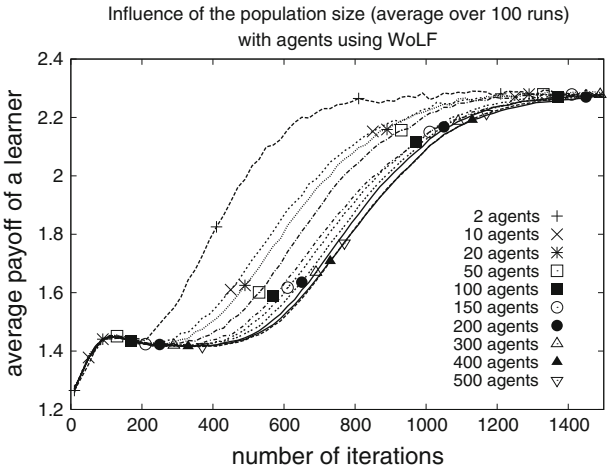

Fig. 3 Dynamics of the average payoff of learners using WoLF with different population sizes (average over 100 runs)

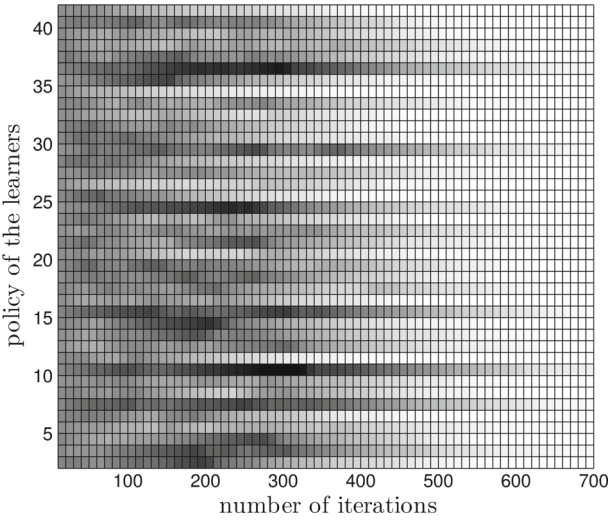

Fig. 4 Dynamics of the probability to play R for each agent (each agent is represented by *two lines*: policy to play as a row and a column player): the *clearer* the cell, the more likely to play L, the *darker*, the more likely to play R. The convention of choosing action L emerges through social learning

graph suggests that after a certain population size, meeting an unknown agent or an agent that was met long time before does not make such a difference on the speed of emergence.

For the rest of paper, we use the coordination game presented in Table 3. This stage game models the situation where agents need to agree on one of several equally desirable alternatives. For example, for the two-action case, this game can represent the situation where agents choose on what side of the road to drive. When both agents drive on their left (action L), or on their right (action R), there is no collision, else there is a penalty. The societal conventions that we would want to evolve are either driving on the left or driving on the right. In Fig. 4, we show the learning dynamics in a population of 20 agents. On the x-axis is the number of

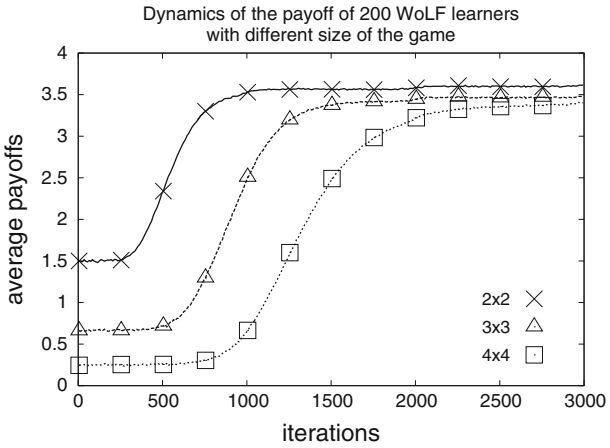

Fig. 5 Dynamics of the payoff of learners using WoLF with different game sizes (average over 100 runs)

iterations. Each agent is represented by two consecutive lines, one for learning to play as a row player and the other to learn to play as a column one. The level of gray in a cell determines the probability of playing action R: the darker it gets, the closer the probability is to 1. In this particular run, we know that the norm where all agents play L emerges. We were expecting a gradual convergence in which the policy of all agents would converge to action L at a similar rate, i.e., in the figure, we expected to observe a gradual change of color from grey to white. Surprisingly, we observed that some agents first get a strong preference for the other action R (we observe very dark cells) but later changed and converged to white. From the early iterations, it is difficult to foresee which action will finally emerge as the convention in the society.

Now, we consider the effect of the number of actions available to each agent. The stylized game, representing other, non-driving scenarios, can be expanded to n-actions: the agents receive a payoff of 4 when they choose the same action and a payoff of -1 when their actions differ. We ran this experiment with $n \in \{2, 3, 4\}$ in a population of 200 agents using WoLF. The results are presented in Fig. 5. When the number of actions increase, the proportion of joint actions with high payoff decreases. When the agents explore at the beginning, the expected utility is less with a larger game. Over time a convention does emerge, with the average payoff of the population approaching 4. It takes longer for a convention to evolve for larger action sets as the space of joint actions increases quadratically.

Finally, we present the effect of the learning algorithm used by the agents in Fig. 6. Since there is no clear choice of learning algorithms to use in general, we wanted to evaluate a few representative learning algorithms. We study the influence of the learning algorithms on a population of 200 agents playing the two-action game. When the entire population uses the same learning algorithm, a convention emerges quicker with a population of Q-Learners ($\approx$ 100 iterations), followed by a population of WoLF ($\approx 1,000$ iterations), and the population of agents using FP ($\approx 40,000$ iterations). The payoff reached at convergence is different for different algorithms due to different exploration schemes. We also show results of hybrid population using equal proportions of any two or all three of these algorithms. The time taken by mixed groups to evolve conventions are in between the time taken by the corresponding homogeneous groups.

We conclude that, even if the "opponent" may (1) be different, (2) have a different history of interaction, (3) have a different learning algorithm, a convention still consistently emerges

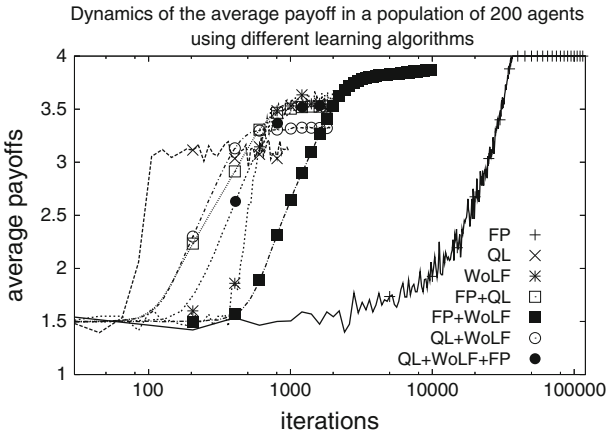

Fig. 6 Dynamics of the payoff of learners using different learning algorithms (population of 200 agents, average over 100 runs)

through social learning. In the following, we study more scenarios (when some agents have a fixed strategy, when agents can only interact with their neighbours in a social network) and we will provide results with a single learning algorithm. We obtained similar results when agents use different learning algorithms, the emergence is simply delayed when slower learning algorithms are used.

4.3 Influence of fixed agents

So far, we have observed that all conventions with equal payoffs were evolved roughly with the same frequency over multiple runs. This is understandable because the payoff matrix in Table 3 does not support any preference for one convention over the other. Extraneous effects, however, can bias a society of learners towards a particular convention. For example, some agents may not have learning capabilities and repeat a pre-determined action. We study the influence of agents playing a fixed pure strategy on the emergence of a convention. For this study, we use the coordination game of Table 3 and consider a population with 3,000 learners, $n_f = 30$ agents playing the fixed strategy 0 (driving on the left), and n_f agents playing strategy 1 (driving on the right). We ran experiments where we add additional agents playing the pure strategy 1. Figure 7 presents the percentage of time that the convention (0, 0), i.e., everyone driving on the left, and (1, 1), i.e., everyone driving on the right, emerges. Note that when there are equal number of agents playing fixed strategy 0 and fixed strategy 1, one of the two conventions emerges with almost equal frequency. It is interesting to note that with only 4 additional agents choosing to drive on the right, the entire population of 3,000 agents almost always converges to driving on the right! There might therefore be some truth to the adage that most fashion trends are decided by a handful of trend setters in Paris!

4.4 Emergence of conventions in social networks

Now we consider agents situated in more restrictive interaction topologies. Each agent is represented by a node in the network and the links represent the possibility of interaction between nodes (or agents). In this section, we consider that agents form a *one-dimensional*

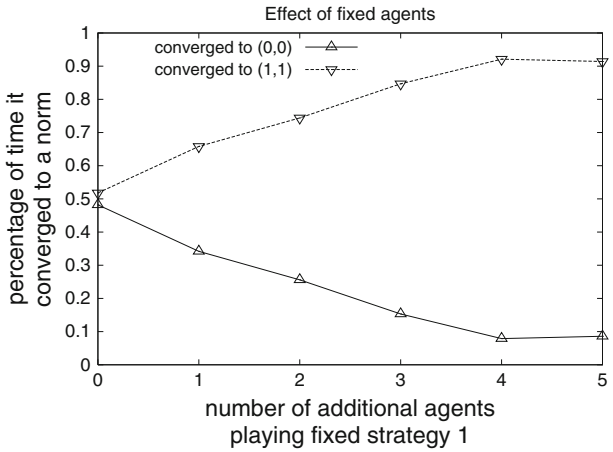

Fig. 7 Number of times each convention emerges (average over 100 runs): a small imbalance in the number of agents using a pure strategy is enough to influence an entire population

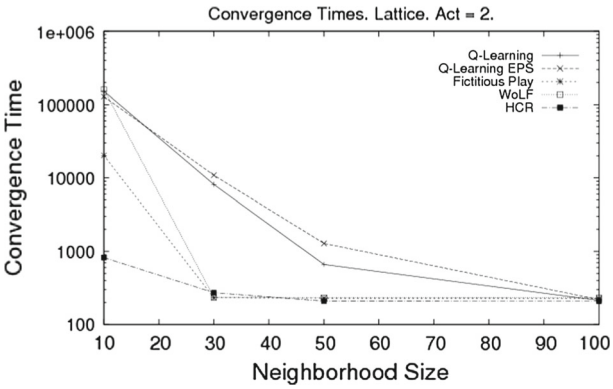

Fig. 8 Convergence times for different neighborhood sizes and different learning algorithms: Q-learning in the figure refers to the classical Q-learning with a fixed exploration rate of 25 %. Q-learning EPS refers to the defined Q-learning in Sect. 3.5 with epsilon-greedy exploration. HCR refers to highest cumulative reward presented in other works ([12,24]), HCR is a deterministic scheme that uses finite memory of size M and chooses the action that fetched the maximum cumulative value over the last M interactions

lattice with connections between all neighboring vertex pairs (example in Figs. 1b). During this experiment, we limited the execution of the simulations to one million timesteps. Figure 8 shows a comparison of convergence times for different neighborhood sizes in a one-dimensional lattice, with different learning algorithms.

We can see that when increasing the neighborhood size, the convergence time is steadily reduced. This effect is due to the topology of the network. When the one dimensional lattice has a small neighborhood size, on average, the diameter of the graph[5] is high and therefore agents located in different parts of the network need a higher number of interactions to communicate their decisions or arrive at a consensus.

[5] The diameter of a graph is the largest number of vertices which must be traversed in order to travel from one vertex to another.

When agents have a small neighborhood size, they will interact often with their neighbors, resulting in diverse subconventions forming at different regions of the network. We note that in each interaction, both agents are learning from it, therefore agents reinforce each other in each interaction. Such divergent subconventions conflict in overlapping regions. To resolve these conflicts, more interactions are needed between agents in the overlap area between regions adopting conflicting subconventions. Unfortunately, the agents in the overlapping regions may have more connections in their own subconvention region and hence will be reinforced more often by their subconventions, which makes it harder to break subconventions to arrive at a consistent, uniform convention over the entire society. On the other hand, when neighborhood sizes are large, and hence network diameters are small, agents interact with a large portion of the population. As a result, it is less likely that subconventions are created or sustained.

5 Emergence of different sub-conventions

Though it is preferable that a convention emerges, it may not always be the case that a single convention emerge in the population. In the following, we provide some evidence showing that sub-conventions can emerge and be stable, and then analyze and explain the cause underlying this phenomenons. First, we model two, completely-connected populations that interact infrequently, and we see that a convention may not emerge when the likleihood of interaction between members of the two populations is below a threshold value. We also observed the existence of multiple sub-conventions in some scale-free networks.

5.1 Emergence of conventions in isolated subpopulations

It is well-documented that isolated populations in seggregated societies can be using contradictory conventions, e.g., driving on the "right" or the "wrong" side of the road. We wanted to replicate this phenomenon using our social learning framework. When two groups of agents interact only infrequently, it is possible that a different convention emerges in each group. In particular, we are interested in studying the degree of isolation required for divergent conventions to emerge in different groups. For our experiments, we consider two groups of equal size and a probability p that agents of different groups interact.

Results from this set of experiments are presented in Fig. 9. We observe that when the probability of interaction is at least 0.3, a single convention pervades the entire population. In roughly half of the runs all agents learn to drive on the left and for the other half they learn to drive on the right. But for interaction probabilities of 0.2 and less, there are runs where divergent conventions emerge in the two groups (corresponding to the white space above the shaded bars in Fig. 9). This is a very interesting observation and we are surprised by the relatively high interaction probabilities that could still sustain divergent conventions.

In the simulations, agents always have a small probability to explore, and they may pick an action at random instead of following the policy learnt. From the point of view of the learning algorithm, this is considered as some noise and does not trigger a sufficient change in the policy. When the probability of interaction between the two groups is high, agents will observe many interactions that does not follow their sub-convention, which is enough to force the emergence of a single convention. However, the simulation suggests that with a low probability of interaction, the presence of a different group using a different sub-convention does not have a sufficient impact as it may be similar to some high level of exploration.

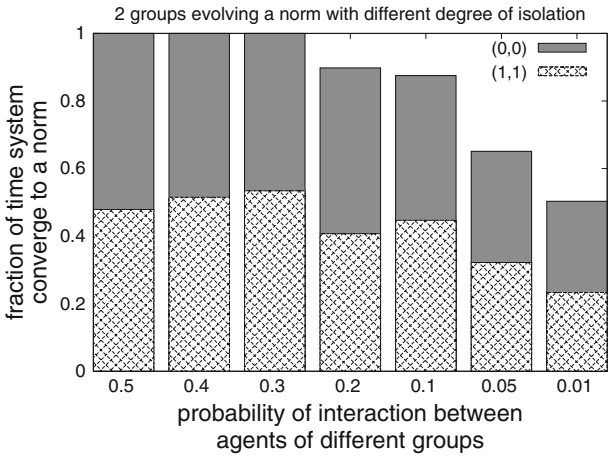

Fig. 9 Two groups of 100 agents each evolve conventions with different interactions frequencies (average over 1,000 runs). When the probability of interaction is low, the groups can evolve different conventions

5.2 Convergence in scale-free networks

We observe another interesting phenomenon for scale-free networks: sub-conventions might be persistent and the entire population fails to converge to a single convention! The coexistence of stable sub-conventions was identified in [39] for the first time in all of our research on convention emergence when using a specific utility function. We performed some additional experiments using our settings and we observed similar results. Contrary to what was believed in previous work, we show that a norm may not always emerge. In this section, we show that the emergence of two stable sub-conventions can be explained by some particular structure of the network.

The explanation of this rather interesting phenomena can be found in some possible substructure of such a network. By paying special attention to the scale-free networks when sub-conventions remain stable, we identify a certain type of structure. We present, in Fig. 10, a portion of a potential scale-free network that represents what we have identified as the sub-convention incubator. We see two subgraphs connected through one central bridging node. On one side, the bridging node is connected to k leaf nodes. On the other side, the bridging node is connected to n hub nodes, themselves connected to many other nodes (which can be connected with each other). Depending on the ratio between the k leaves and the n hubs, it is possible that a different convention emerges on the k leaf nodes and the n hub nodes. This is a special case of the previous experiments in which two populations are communicating infrequently. Here, the two populations only interact through one node: the central node.

The method used in this work for the generation of the scale-free networks is the classic *Preferential Attachment* [5]. With this method the network is constructed progressively, adding nodes to an initial single-node network, and with a probability of linking nodes proportional to the number of existing links of each node. In other words, better connected nodes have higher chances to be connected with newly created nodes. However, and as it is a probabilistic process, the rest of nodes can be connected to newly created nodes. The resulting network has therefore a degree distribution that follows a power-law distribution, meaning that some nodes have a high degree, forming a hub of nodes, when others have small ones. Moreover, and inherent to the scale-free network, the clustering coefficient distribution

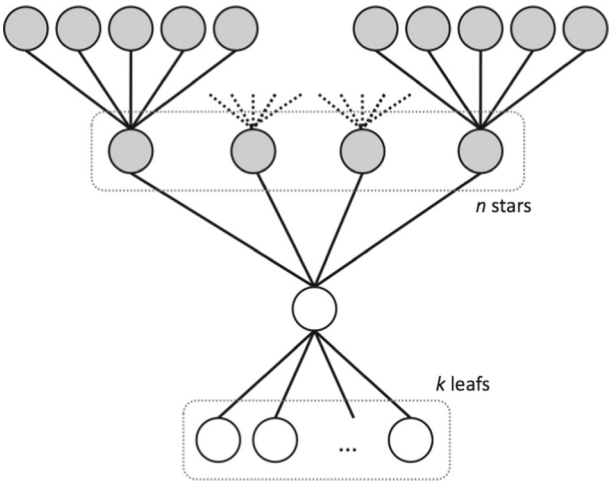

Fig. 10 *k–n* Connected stars network

decreases as the node degree increases and also follows a power law. This means that low-degree nodes are connected to dense sub-graphs and that those sub-graphs are connected to each other through hubs. In other words, there are different communities, and few agents are part of two or more communities. It is then possible to have a community that has very few members connected with other communities at the edge of the network. Hence, we find ourselves in a similar situation as in the previous section with two groups of agents interacting unfrequently, allowing sub-conventions to remain meta-stable within those communities.

One possible line of future research is to study more closely the graph-structure in which a sub-convention may emerge. Though we have studied some particular examples, we are yet to provide a general description about such structures. The important claim we want to make in this paper is that sub-conventions may arise in scale-free networks and that they are sustained by a group of agents that interacts very infrequently with agents oustide their group.

6 Conclusions

We investigated a bottom-up process for the emergence of social convention that depends exclusively on individual experiences rather than observations or hearsay. Our proposed social learning framework requires each agent to learn from repeated interactions for a given social situation, without using knowledge of the identity of the other agents involved in the interactions. The goal of this work was to evaluate whether such social learning can successfully evolve and sustain a useful social convention that resolves conflicts and facilitates coordination between population members. Our experimental results confirm that such distributed, individual, and social (interacting with many individuals rather than repeated interactions with the same agent) learning is indeed a robust mechanism for evolving stable social conventions. Our results suggest that private interactions are sufficient for a convention to emerge. Additional information (e.g. observation of interactions of other agents) can improve the speed of emergence, but it is not required. Our results also suggest that to deploy a multiagent system, one can use generic agents that use learning mechanisms which can,

with no, or very little, detailed knowledge of the environment, learn efficient and stable coordination behavior.

We investigate the effects of population size, number of actions, different learning strategies, non-learning agents, and multiple relatively isolated populations on the speed and stability of convention evolution. We confirmed that even thorny problems like social dilemmas can be successfully addressed by the social learning framework. This is quite encouraging as the agents are learning from data gathered from interactions with many different agents, and not a single opponent as is usually assumed.

We have identified certain challenges to the emergence and maintenance of subconventions in particular types of topologies and mainly in scale-free networks. This was a surprise as previous research demonstrated consistent emergence of a convention in all cases. We hypothesize that stable subconventions, preventing emergence of globally consistent conventions, arise in scale-free networks because of some inherent structural characteristics of these networks. We plan to investigate, in further depth, the reasons why these subconventions might be created and maintained, as well as, mechanisms to dissolve them.

For our simulations, we have assumed that each agent recognizes a given social situation and was adapting his behavior for that particular situation. In real application, an agent will need to perform this recognition, which may not always be easy. In particular, if an agent realizes that no behavior is satisfying, she will need to define sub-situations or exceptions. It will be interesting to study mechanisms to learn this efficiently.

We would like to study other intriguing phenomena like punctuated equilibria in social convention evolution [46] within our framework. Other interesting experiments include study of spatial distribution of agents and corresponding effects on rate and divergence of convention emergence.

References

1. Airiau, S., Saha, S., & Sen, S. (2007). Evolutionary tournament-based comparison of learning and non-learning algorithms for iterated games. *Journal of Artificial Societies and Social Simulation*, *10*(3), 1–7.
2. Axelrod, R. (1986). An evolutionary approach to norms. *The American Political Science Review*, *80*(4), 1095–1111.
3. Axelrod, R. (1997). *The complexity of cooperation: Agent-based models of conflict and cooperation*. Princeton, NJ: Princeton University Press.
4. Banerjee, D., & Sen, S. (2007). Reaching pareto-optimality in prisoner's dilemma using conditional joint action learning. *Autonomous Agents and Multi-Agent Systems*, *15*, 91–108.
5. Barabási, A., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–512.
6. Boella, G., & Lesmo, L. (2002). A game theoretic approach to norms and agents. *Cognitive Science Quarterly*, *2*(2–3), 492–512.
7. Boella, G., & van der Torre, L. (2003). Norm governed multiagent systems: The delegation of control to autonomous agents. In *IAT '03: Proceedings of the IEEE/WIC International Conference on Intelligent Agent Technology* (p. 329). Washington, DC, USA: IEEE Computer Society.
8. Bowling, M. (2005). Convergence and no-regret in multiagent learning. *Advances in Neural Information Processing Systems*, *17*, 209–216.
9. Bowling, M. H., & Veloso, M. M. (2002). Multiagent learning using a variable learning rate. *Artificial Intelligence*, *136*(2), 215–250.
10. Coleman, J. S. (1998). *Foundations of social theory*. Cambridge, MA: Harvard University Press.
11. Crandall, J. W., & Goodrich, M. A. (2005). Learning to compete, compromise, and cooperate in repeated general-sum games. In *Proceedings of the 22nd International Conference on, Machine Learning* (pp. 161–168).
12. Delgado, J. (2002). Emergence of social conventions in complex networks. *Artificial Intelligence*, *141*(1–2), 171–185.
13. Dignum, F., Kinny, D., & Sonenberg, L. (2002). From desires, obligations and norms to goals. *Cognitive Science Quarterly*, *2*(3–4), 407–430.

14. Epstein, J. M. (2001). Learning to be thoughtless: Social norms and individual computation. *Computational Economics*, *18*(1), 9–24.
15. Fudenberg, D., & Levine, D. K. (1998). *The theory of learning in games*. Cambridge, MA: MIT Press.
16. García-Camino, A., Rodríguez-Aguilar, J. A., Sierra, C., & Vasconcelos, W. (2006). A rule-based approach to norm-oriented programming of electronic institutions. *SIGecom Exchanges*, *5*(5), 33–40.
17. Genesereth, M. R., Ginsberg, M. L., & Rosenschein, J. S. (1988). Cooperation without communication. *Distributed Artificial Intelligence*, 220–226.
18. Hales, D., & Arteconi, S. (2006). Slacer: A self-organizing protocol for coordination in p2p networks. *IEEE Intelligent Systems*, *21*(29), 35.
19. Kandori, M., & Rob, R. (1995). Evolution of equilibria in the long run: A general theory and applications. *Journal of Economic Theory*, *65*(2), 383–414.
20. Kittock, J. E. (1995). Emergent conventions and the structure of multi-agent systems. In *Proceedings of the 1993 Santa Fe Institute Complex Systems Summer School. Santa Fe Institute Studies in the Sciences of Complexity Lecture*. Santa Fe Institute.
21. Lakkaraju, K., & Gasser, L. (2006). Population and agent based models for language convergence. In *Proceedings of the 21st National Conference on Artificial Intelligence, AAAI'06* (Vol. 2, pp. 1887–1888). Menlo Park: AAAI Press.
22. Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.
23. Montanari, A., & Saberi, A. (2010). The spread of innovations in social networks. *Proceedings of the National Academy of Sciences*, *107*(47), 20196–20201.
24. Mukherjee, P., Airiau, S., & Sen, S. (2008). Norm emergence under constrained interactions in diverse societies. In *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS-08)* (pp. 779–786). Estoril, Portugal.
25. Myerson, R. B. (1991). *Game theory: Analysis of conflict*. Cambridge, MA: Harvard University Press.
26. Nowak, M. A., & May, R. M. (1992). Evolutionary games and spatial chaos. *Nature*, *359*, 826–829.
27. Nudelman, E., Wortman, J., Shoham, Y., & Leyton-Brown, K. (2004). Run the gamut: A comprehensive approach to evaluating game-theoretic algorithms. In S. A (Ed.), *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '04)* (pp. 880–887). Washington, DC, USA: IEEE Computer Society.
28. Panait, L., & Luke, S. (2005). Cooperative multi-agent learning: The state of the art. *Autonomous Agents and Multi-Agent Systems*, *11*(3), 387–434.
29. Powers, R., Shoham, Y., & Vu, T. (2007). A general criterion and an algorithmic framework for learning in multi-agent systems. *Machine Learning*, *67*(1–2), 45–76.
30. Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
31. Schweitzer, F., Zimmermann, J., & Mühlenbein, H. (2002). Coordination of decisions in a spatial agent model. *Physica A*, *303*(1–2), 189–216.
32. Sen, S., & Airiau, S. (2007). Emergence of norms through social learning. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI'07)* (pp. 1057–1512).
33. Shoham, Y., & Tennenholtz, M. (1997). On the emergence of social conventions: Modeling, analysis, and simulations. *Artificial Intelligence*, *94*(1–2), 139–166.
34. Tinnemeier, N., Dastani, M., & Meyer, J.-J. (2010). Programming norm change. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '10)* (Vol. 1, pp. 957–964). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.
35. Tuyls, K., & Nowé, A. (2005). Evolutionary game theory and multi-agent reinforcement learning. *Knowledge Engineering Review*, *20*(1), 63–90.
36. Urbano, P., Balsa, J., Antunes, L., & Moniz, L. (2009). Force versus majority: A comparison in convention emergence efficiency. In *Coordination, Organizations, Institutions and Norms in Agent Systems IV: COIN 2008 International Workshops, COIN@AAMAS 2008, Estoril, Portugal, May 12, 2008* (pp. 48–63). Berlin: Springer.
37. Vázquez-Salceda, J., Aldewereld, H., & Dignum, F. P. M. (2005). Norms in multiagent systems: From theory to practice. *International Journal of Computer Systems Science & Engineering*, *20*(4), 225–236.
38. Verhagen, H. (2001). Simulation of the learning of norms. *Social Science Computer Review*, *19*(3), 296–306.
39. Villatoro, D., Sen, S., & Sabater-Mir, J. (2009). Topology and memory effect on convention emergence. In S. A (Ed.), *Proceedings of the International Conference of Intelligent Agent Technology (IAT)*. Hoboken, NJ: IEEE Press.
40. Villatoro, D., Sen, S., & Sabater-Mir, J. (2010). Of social norms and sanctioning: A game theoretical overview. *International Journal of Agent Technologies and Systems*, *2*, 1–15.

41. Vu, T., Powers, R., & Shoham, Y. (2006). Learning against multiple opponents. In *AAMAS '06: Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems* (pp. 752–759). New York, NY, USA: ACM.
42. Walker, A., & Wooldridge, M. (1995). Understanding the emergence of conventions in multi-agent systems. In *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS95)* (pp. 384–389). San Francisco.
43. Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3–4), 279–292.
44. Wolpert, D. H., & Tumer, K. (2002). Collective intelligence, data routing and Braess' paradox. *Journal of Artificial Intelligence Research*, *16*, 359–387.
45. Young, P. H. (1993). The evolution of conventions. *Econometrica*, *61*(1), 57–84.
46. Young, P. H. (1996). The economics of convention. *The Journal of Economic Perspectives*, *10*(2), 105–122.