Auton Agent Multi-Agent Syst
DOI 10.1007/s10458-012-9193-x

Manipulating convention emergence using influencer agents

Henry Franks · Nathan Griffiths · Arshad Jhumka

Abstract Coordination in open multi-agent systems (MAS) can reduce costs to agents associated with conflicting goals and actions, allowing artificial societies to attain higher levels of aggregate utility. Techniques for increasing coordination typically involve incorporating notions of *conventions*, namely socially adopted standards of behaviour, at either an agent or system level. As system designers cannot necessarily create high quality conventions a priori, we require an understanding of how agents can dynamically generate, adopt and adapt conventions during their normal interaction processes. Many open MAS domains, such as peer-to-peer and mobile ad-hoc networks, exhibit properties that restrict the application of the mechanisms that are often used, especially those requiring the incorporation of additional components at an agent or society level. In this paper, we use *Influencer Agents* (IAs) to manipulate convention emergence, which we define as agents with strategies and goals chosen to aid the emergence of high quality conventions in domains characterised by heterogeneous ownership and uniform levels of agent authority. Using the *language coordination* problem (Steels in Artif Life 2(3):319–392, 1995), we evaluate the effect of IAs on convention emergence in a population. We show that relatively low proportions of IAs can (i) effectively manipulate the emergence of high-quality conventions, and (ii) increase convention adoption and quality. We make no assumptions involving agent mechanism design or internal architecture beyond the usual assumption of rationality. Our results demonstrate the fragility of convention emergence in the presence of malicious or faulty agents that attempt to propagate low quality conventions, and confirm the importance of social network structure in convention adoption.

H. Franks (✉) · N. Griffiths · A. Jhumka
Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK
e-mail: hpwfranks@googlemail.com

N. Griffiths
e-mail: N.E.Griffiths@warwick.ac.uk

A. Jhumka
e-mail: H.A.Jhumka@warwick.ac.uk

Published online: 15 February 2012

Keywords Conventions · Norms · Agent coordination · Convention emergence · Social influence

1 Introduction

Coordination is fundamental to the effectiveness and coherence of systems in which there is no centralised control. Open multi-agent system (MAS) domains such as peer-to-peer (P2P) and mobile ad-hoc networks (MANETs) exhibit a number of properties that make coordination difficult: (i) decentralised control with all agents having equal authority, (ii) heterogeneous ownership, (iii) limited resources, and (iv) complex connecting topologies. Given the existence of decentralised control and uniform authority, traditional approaches have involved either protective mechanisms, such as trust and reputation [15,28,29,42] or the use of notions of aggregate behaviour such as *conventions* and *norms* [31,35,43].

Conventions, which are socially adopted standards of behaviour, are known to encourage high levels of coordination, but efficiently and robustly facilitating the emergence of high-quality conventions remains an open research problem. Considerations of limited knowledge of society characteristics, time variance, and computational difficulty often preclude the ability to generate high-quality conventions a priori. Mechanisms that aid on-line generation and adoption of appropriate conventions often assume the ability to universally incorporate additional structures into agent or society architecture. By adopting the property of heterogeneous ownership of agents ((ii) above) and assuming (as with all open systems) that agents can join and leave freely, we cannot rely on the ability to add additional structures into agent architectures, or make any guarantees that the proportion of agents adopting a particular mechanism will be sufficient to ensure feasibility. Similarly, we cannot assume that we can impose society-level structures on the system. As such, we require a model of how purely rational agents might be manipulated into adopting high-quality conventions and otherwise aided in increasing levels of coordination within the system.

In this paper, we propose inserting a number of agents, with specific conventions and strategies, such that the agent population as a whole, through their normal rational selection of actions, is guided towards the adoption of high-quality conventions. We call these inserted agents *Influencer Agents* (IAs), and show that a small proportion of IAs in an artificial society can efficiently aid the generation and propagation of high quality conventions. This mechanism of manipulating convention emergence does not require any assumptions of agent behaviour beyond rationality, and to our knowledge does not currently exist in this form in the literature.

We discuss current approaches to coordination and convention emergence, along with our adopted model for exploring convention emergence and inspirations for the IA concept, in Section 2. Section 3 details our experimental setup. Section 4 establishes baseline model behaviour and presents data demonstrating the efficacy of the IA concept and other results. Finally, Section 5 discusses conclusions and directions for future work.

2 Background

We begin this section by introducing the notion of conventions and a selection of the mechanisms that have been proposed for achieving convention emergence. We then discuss the various techniques that act as inspiration for our IA mechanism. Finally, we describe the language coordination problem that we use for the evaluation of our proposed IA mechanism.

2.1 Conventions

Conventions notionally represent socially-accepted rules or standards of behaviour. There is no obligation to act in a certain way but instead there exists an expectation that to not act in that way will result in costs to the agent and the society through reduced coordination. Conventions are often considered as emergent from the aggregation of local interactions and information [18,37,43], and are a powerful abstraction tool for modelling the complex interactions between self-interested individuals. The use of conventions in artificial societies greatly increases coordination [16], reducing conflict in the goals and actions of agents. As such, conventions provide useful constraints on the set of possible actions [6,37]. Given the rise in open MAS domains whose features preclude the use of pre-generated conventions, we consider conventions which agents generate and adopt online without the need for a centralised authority. How this can be done efficiently, effectively and robustly remains an open research problem.

It is important to distinguish between conventions and *norms*, which also represent socially-accepted rules governing behaviour, but are generally considered to include an obligation to act according to the norm. Norms are thus a stronger form of convention, with mechanisms to encourage norm emergence typically including *incentives* or *sanctions* to motivate agent adherence [2,4,21,27,34,36]. Such mechanisms require additional agent-level or society-level components, which may not always be practical. In this paper, we focus on conventions as a softer form of societal rule as we are (i) concerned with societies in which we make as few assumptions as possible about agent mechanisms and behaviour, and (ii) motivated by influencing agent actions such that costs associated with malcoordination are reduced, without obliging agents to act in a certain way.

Mechanisms for distributed convention emergence can be broadly categorised into (i) those that necessarily require incorporation of additional components in either agent or society architecture, and (ii) those that rely on rational action selection to create aggregate consensus with less intrusive assumptions about agent behaviour. For example, Salazar et al. [30] use a spreading mechanism incorporating several components: information transfer, selection, noise, innovation and self-protection. The authors report promising results, with fast convergence to high quality conventions using realistic assumptions (e.g. a large convention space and complex connecting topologies), but require several additional components to be built into agents and universal adoption of the mechanism by all agents. The seminal exploration of norm emergence introduced by Axelrod [4] requires that agents incorporate punishment (at personal cost) into their strategies, and Grizard et al. [14] link reputation assessments with convention adherence, but require a system-wide monitoring mechanism to be built into the society. Much of this and other literature (e.g. [18]) reports encouraging results but is often restricted by limiting assumptions in the model (e.g. a lack of connecting topology in the work of Axelrod [4], or intrusive architectural requirements in the approaches of Grizard et al. [14] and Salazar et al. [30]). In this work, we aim to have fewer limiting assumptions and to model features of realistic domains such as complex connecting topologies, but we note that we do still make certain simplifying assumptions: our topologies are static, and our agents, with the exception of our IA agents, are homogeneous (see Sect. 2.4).

A number of mechanisms for convention emergence based on rational action selection have been investigated. For example, Sen and Airiau [35] describe convention emergence through social learning, in which each agent learns (using a Q-learning algorithm) the best action through private interactions. As discussed below, their contributions are highly relevant and demonstrate the feasibility of convention emergence through local interactions without complex additional agent architecture, but again are limited by several simplifying

assumptions (including a lack of connecting topology and a convention space with only two possible conventions). Shoham and Tennenholtz [37] formalise conventions in a game theoretic setting and assume that agents continually evaluate personal strategies given new information. Their contributions provide strong analytical insight into convention emergence but the results are given in the context of a highly abstracted model. Walker and Wooldridge [43] investigate a formal model of convention emergence in which agents choose from among a set of strategies based on observation of other agents' behaviours. Their contribution is one of the earliest in the area, and presents tentative results demonstrating the emergence of conventions at a societal level from local interactions with minimal assumptions about additional agent architecture.

Existing approaches make several simplifying assumptions about agent behaviour involving the ability to universally include additional components in agent architectures. As noted above, we are concerned with domains in which disparately owned agents can join and leave freely, and thus the only safe assumption we can make is that of rationality. In this context, we interpret this such that agents will select strategies that minimise costs associated with miscoordination. In a real-world setting, agents are likely to include more advanced behavioural models, such as notions of trust or other mechanisms for protection against malicious agents. One of our motivations for attempting to minimise assumptions about agent behaviour is that we cannot know how these systems will be implemented, their parameters, or the levels of heterogeneity between agents implementing them. Restricting our base assumptions allows us to explore society dynamics in a more general setting.

2.2 Influencer agents

We define an influencer agent (IA) as an agent inserted by any interested party (typically the system designer or manager) with the specific goal of influencing and aiding the emergence of appropriate conventions, for example to increase the aggregate utility of an artificial society. In this paper we are concerned with facilitating the *emergence of a single, high quality convention*, but IAs might also be used to block the emergence of certain conventions or coordinate the emergence of multiple appropriate conventions (such as in the El-Farol Bar problem [3]). This is further discussed in Sect. 4.6. IAs were inspired by a number of contributions in the literature that include the notion of a small proportion of unprivileged agents influencing the aggregate behaviour of an entire artificial society.

Our initial inspiration is based on work by Garlick and Chli [11], who created an agent-based model to investigate the effects of curfews in civil disturbances. While their domain of interest is very different to ours, Garlick and Chli consider two important concepts: (i) a small proportion of policemen agents attempting to influence the society towards peaceful outcomes, and (ii) a notion of social influence based on communication between agents. They found that restricting communication, and thus influence, could significantly change the outcome of the model, and that free communication allowed agents to direct large populations towards their preferred outcome (i.e. rebellion or peace). Given that we expect few limitations on communication in our domains beyond the usual factors of noise and node failure, this may translate to realistic open MAS domains.

While exploring how agents with fixed convention adherence (i.e. that use one strategy without possibility of changing) affect the conventions that emerge, Sen and Airiau [35] found that 4 agents fixed on one strategy (of two alternatives) was sufficient to influence a population of 3000 agents to adopt that strategy. These results suggest that small numbers of agents can heavily influence large groups of self-interested individuals. However, Sen and Airiau's model is limited by three assumptions: (i) there are only two possible conventions,

(ii) there is no topology defining the communication links between agents (instead, agents are randomly paired from the whole population), and (iii) interactions are private. With the intention of moving the model towards more realistic settings, we adopt a domain with many potential conventions (10^{10} with the parameter settings used in Sect. 4 where we discuss our results) and situate agents within a connecting topology.

Yu et al. [44] show that small sets of informed individuals can guide large groups towards coordinated outcomes, with the aim of solving problems of distributed consensus. However, their approach requires significant additional components of agent architecture. Similarly, Oh and Smith [24] discuss using a subset of agents in a population as *leaders* for other agents in multi-agent learning for resource allocation problems, such that the agents who follow them are saved the computational burden and other costs of convention generation. The authors argue that this approach aids convention emergence in highly dynamic societies since new agents can employ social learning rather than environmental exploration when entering the system. Our contribution has some similarities here, in that IAs are analogous to leader agents, and similarly bear costs associated with convention emergence (although in our current formulation IAs do not generate conventions, this being an area for future investigation).

Axelrod's model of norm emergence [4] requires observing agents to punish norm violators, and results in high levels of emergent cooperation. However, the model considers populations in which the entire population is able to punish norm violators, agents are not situated on a connecting topology, and the convention space is limited to two dimensions.

Grizard et al. [14] consider a system which links reputation assessments with cooperative social norms, where control agents are injected to monitor agents and sanction their behaviour (if necessary) by reducing their reputation, which leads to ostracism effects. They obtain encouraging results, but require the imposition of society level components. Despite homogeneous authority and large populations, individual agents can clearly have a significant effect on emergent social dynamics.

Little work has been done on the generation of conventions themselves. Recently, Morales et al. [22] have presented work on generating conventions using historical data on the success of a given convention. They situate agents in an abstract traffic model and use monitoring agents to determine the efficacy of imposed conventions. A machine-learning algorithm generates new conventions as necessary and these are communicated to the agents in the environment. Their work is one of few to address the generation of norms and conventions and there are parallels between their monitoring agents and our IAs. However, their model requires a central authority to process convention data and generate new conventions, whereas our IAs act independently and attempt to influence nearby agents to a given convention.

We consider systems in which all agents are given equal authority, and thus we cannot elevate the privileges or abilities of any inserted IAs above the rest of the population. However, there is still potential for significant influence. We can identify a number of potential strategies IAs might use: (i) lead-by-example, (ii) incentives, (iii) sanctions, and (iv) information propagation. We present an overview of inspirations for these potential strategies, but in the remainder of the paper we are only concerned with leading-by-example as a means to explore the feasibility of the IA concept.

Lead-by-Example: Sen and Airiau [35] use a model of private interactions that introduced the notion of small sets of agents with fixed strategies being able to affect the norms adopted in a relatively large population. This implies that IAs may be able to choose strategies that enable them to *lead-by-example*, interacting with other agents using actions determined by high quality conventions. Agents observe these actions and incorporate them into their own strategies, allowing the convention to spread throughout the population. Additional targeting

of where to insert high-influence agents (e.g. by using topological information) might further increase the efficacy and robustness of this strategy, although this was not considered by Sen and Airiau.

Incentives and Sanctions: Both *incentives* and *sanctions* have been extensively studied in the literature over a wide variety of fields, and both are known to play a significant role in the emergence and enforcement of social norms. Oliver [25] concludes that incentives are an effective way to motivate small groups, whereas sanctions are more effective at generating unanimous cooperation once a small group has been established (though at the expense of potentially generating hostilities that will disrupt such cooperation). The model of civil violence described by Garlick and Chli [11] sanctions agents by restricting communications, and their results show that this significantly influences the normative outcome.[1] Axelrod [4] showed that punishment for norm violation could create stable cooperative populations, although more recent investigations have cast doubt on the scalability of results from this model [10,20]. Despite this, sanctions and incentives remain a powerful idea for aiding convention and norm emergence. Implementation of sanctions and incentives is likely to be domain specific, and it is not intuitively clear how IAs might be able to effectively incorporate these notions into their strategies.

Information Propagation: It has long been accepted that *information propagation* plays a key role in social dynamics. Garlick and Chli's restriction of communications inherent in imposing a curfew created significant effects. Gossiping of information can replace the need for direct observation of interactions [39], which is a core component of many of the convention emergence models discussed above (e.g. [37,43]). IAs could be used to propagate trust and reputation assessments, high quality conventions, or other useful information, and could even block or otherwise disrupt communications from non-compliant agents.

This paper is concerned only with the strategy of lead-by-example as a demonstration of the feasibility of the notion of IAs. We aim to confirm the hypothesis that small proportions of agents in a population can significantly influence convention emergence in an artificial society. As such, the agents that we consider are as simple as possible in terms of strategy. Specifically, the IAs we propose in this paper are identical to the other agents in the population, except that they do not adapt their behaviour in response to conventions proposed by others (instead, they adhere only to their own, fixed convention). Informally, we view such IAs as attempting to lead-by-example. This use of IAs relies on the rationality of agents: IAs attempt to propagate high-quality conventions, and agents adopting these conventions avoid costs associated with malcoordination.

IAs represent a model in which a small proportion of inflexible agents spread a given convention for the duration of the simulation. We also present a second model, in Sect. 4, in which we give a specific initial convention to a (potentially larger) proportion of agents and then let them continue as normal. These alternatives apply to different potential real-world situations; the latter being a good fit for domains in which we can temporarily influence a large set of agents, and the former being more suited to situations in which we can insert a small number of agents explicitly under our ownership and control. In Sect. 4, we use the latter model in some of our simulations when validating certain aspects of our IA model. However, in this paper we are primarily interested in characterising and quantifying the effects small groups of agents can have on populations many times their size, rather than the effects of groups of flexible agents adopting and adapting an initially implanted convention.

[1] Note that this influence is not always towards the preferred outcome, with the direction the population takes being dependent on its current state.

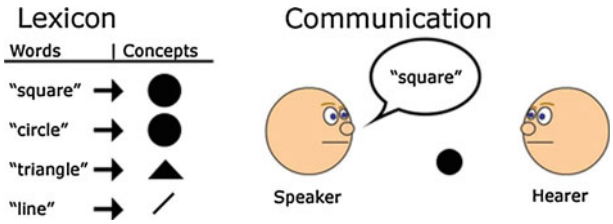

Fig. 1 Illustration of a lexicon and a simple communication action

2.3 Linguistic coordination domain

To illustrate our approach, we require a domain that effectively captures the difficulties of convention emergence. The language coordination problem, initially proposed by Steels [40], is particularly applicable: there is a large number of potential conventions, and there exists a partial ordering on convention quality. We adopt the extended model of this domain as described by Salazar et al. [30].

Each agent is associated with a *lexicon*, which is a set of mappings from *words* to *concepts*. We denote the set of words W and the set of concepts C. We assume that $|W| = |C|$. We denote the set of mappings in a lexicon from elements of W to elements of C as M. We assume, for simplicity, that $|M| = |W|$. Each agent starts with a randomised lexicon, meaning that each element of W is mapped to a randomly selected element of C (such that multiple words may map to the same concept).

Agents attempt to communicate with each other using their lexicons, and track the success of their communications. Furthermore, they also exchange and update their lexicons in a manner analagous to a distributed genetic algorithm. Agreeing on a shared lexicon allows agents to communicate effectively with each other, and reduces costs associated with miscommunication. As such, a shared lexicon represents our notion of a convention. We refer to the (potentially partial) lexicons that agents communicate as *convention seeds*, since they have the potential to become established conventions. Figure 1 illustrates the lexicon structure and agent communications.

In this domain, there are many potential conventions[2] with an intrinsic quality metric called *lexicon specificity*. The specificity of a lexicon is the proportion of words that identify a single concept, such that a one-to-one mapping gives a specificity of 1, while a two-to-one mapping gives 0.5 specificity. To calculate specificity for a single concept in a lexicon, we use the formula $S_c = \frac{1}{|W_c|}$, where S_c is the specificity of concept c and W_c is the set of words that identify that concept. If no words identify a concept then $S_c = 0$. The specificity of a lexicon is defined as the average of the specificity of all concepts, or formally:

$$S = \frac{\sum_{c \in C} S_c}{|C|}$$

There may exist multiple conventions of the same quality, in which case it does not matter which one the agents choose as long as they agree. Adhering to a convention allows an agent to avoid the cost associated with being unable to communicate successfully with others. Given the potential size of lexicons, it is not practical for agents to propagate entire vocabularies. This means that agents have incomplete information about other agents' lexicons, and can thus only estimate their quality. We do not know a priori if an ideal lexicon exists in the

[2] There are w^c possible conventions, where w is the number of words and c is the number of concepts [30].

population. Convention emergence is thus a highly challenging problem in the language domain, making it a useful setting for the investigation of convention emergence dynamics. We note that the results of Salazar et al. [30] show efficient and fast convergence to a high-quality convention, but require extensive additional architecture to be built into agents. In our investigation, we have replicated the core components of the convention spreading mechanism introduced by Salazar et al. [30] of information transfer and selection. We assert that these components will be universally adopted by agents that are rational: agents will choose the best convention they can based on the available information, and attempt to reduce costs associated with malcoordination by spreading their "way of doing things".[3] We use this domain to illustrate the ability of small numbers of unprivileged IAs to influence the behaviour of a population.

2.4 Simplifying assumptions

We make two main simplifying assumptions in this paper, namely that (i) the underlying connecting topology is static, and (ii) our agents are homogeneous (aside from the minor differences of IAs).

Static connecting topologies are known to induce different system dynamics than dynamic topologies, and consequently there are limits to how far we can generalise our model to domains characterised by high levels of churn. However, there is relatively little work on modelling dynamic topologies, and defining how a topology is likely to change over time is likely to involve incorporating domain-specific assumptions about agent or environment behaviour. Such modifications are outside the scope of this paper, although an investigation of the efficacy of our model in the presence of topological dynamism is part of our planned future work (see Sect. 5 for further discussion).

There has been significant work on dealing with agent heterogeneity, but it remains an open research area. Typical approaches involve specifying agent communication languages and protocols, although there is a risk of increasing barriers to participation, particularly for agent societies with disparate levels of agent complexity [8]. Singh [38] argues that the majority of existing communication languages are insufficient for application in open MAS domains, which tend to exhibit high levels of agent autonomy and heterogeneity. Extending our contributions to populations of heterogeneous agents is thus a necessary step for future work, but is beyond the scope of the current paper.

3 Experimental setup

We investigate the ability of IAs within an open MAS to influence convention emergence in a society. Our experimental setup is based on that used by Salazar et al. [30], with some elaboration where details of the original configuration are unspecified. Agents are situated within a communication topology that constrains the selection of neighbours with whom they can communicate. In this paper we focus on small-world and scale-free networks, as these topologies reflect features of realistic domains [1].

Within each timestep of the simulation, there are three phases: (i) agent communication, (ii) lexicon spreading, and (iii) lexicon update. The communication, spreading, and updating actions are split into separate loops to prevent unforeseen synchronisation effects, such as

[3] Spreading of conventions can also be interpreted observationally, such that agents receiving a convention from another can be said to be observing that agent's convention.

agents updating their lexicons from out-of-date information. The full control loop for the simulation is shown in Algorithm 1.

3.1 Communication

In the first phase, every agent is in turn denoted as the *speaker*. The speaker selects a random neighbour, denoted as the *hearer*, and sends them a one-word communication about a concept. It is assumed that the speaker can assess whether the hearer understood the communication or not, such that an agent can calculate their own *communicative efficacy*. Salazar et al. [32] define communicative efficacy as the difference between successful (understood) and unsuccessful (not understood) communications, which is calculated every 20 timesteps. For clarity of analysis we use a normalised form of communicative efficacy, defined as the proportion of communications that were successful over the last 20 timesteps. We discuss this in more detail in Sect. 3.5.

3.2 Spreading

After all agents have acted as a speaker in a timestep, each agent is in turn given a chance to send their lexicon (either partially or completely) to all their neighbours. We assume that agents update their lexicons based on information provided by other agents (i.e., there is no centralised authority). Consideration must be given to the synchronicity of the agent update processes. Given the literature concerning the limitations of synchronous strategy update [26,41], we use an asynchronous probabilistic update model: during each timestep, an agent sends lexicon information (see below) with probability p_{send}, and updates its lexicon based on received information with probability p_{update}. This differs from Salazar et al. [30] who combine both probabilities into a single value for spreading. They do not explicitly state when agents initiate their update processes, or whether they are synchronised, but the model presented in Salazar et al. [33] suggests that agents update after a given number of timesteps with a given probability.

Salazar et al. [32] consider two potential mechanisms for lexicon spreading: *Complete* transfer (also called *Copy* transfer) and *Partial* transfer. In complete transfer, agents send their entire lexicons to their neighbours. This should only be seen as an idealised scenario—in practice, it is likely that resource constraints such as bandwidth will mean that only partial transfer is possible, and we therefore focus on partial transfer in this paper. Salazar et al. [32] state that their partial transfer mechanism is based on recombination techniques from evolutionary algorithms literature, but they give no further detail.

We define partial transfer using a two point crossover mechanism that mirrors two-point crossover in genetic algorithms, in which two points are selected in the parent gene strings. Everything between those two points is swapped, generating two offspring. In our case, we generate a single offspring, in the form of a new lexicon, using the following mechanism:

1. Each agent is associated with an integer l, individually chosen uniformly at random at the start of the simulation, that defines the lexicon transfer length. This represents the number of mappings an agent will try to spread from its lexicon to other agents (i.e. $l = |W|$ corresponds to complete transfer). We assume, to simplify our model, that all agents have the same number of words and concepts in their lexicons.
2. When an agent decides to initiate a partial transfer, it selects a random start point, a, in its lexicon such that $a + l \leq |W|$.
3. Then, l mappings are selected from the lexicon starting at mapping a, and are communicated to the recipient(s), which are all neighbours of the sender.

4. If a recipient chooses to incorporate these mappings in the update phase then it replaces l mappings in its own lexicon, starting at a, with the received mappings.

Algorithm 1 Simulation loop

```
for ticks = 1 to simulationLength do

  //Communication Phase
  for all Agent ∈ Population do
    Speaker ← Agent
    Hearer ← getRandomNeighbour(Speaker)
    outcome ← Hearer.hear(Speaker.speak())
    Speaker.updateSuccess(outcome)
  end for

  //Spreading
  for all Agent ∈ Population do
    r ← randomDouble()
    if r < pSend then
      Agent.sendLexicon(Agent.getNeighbours())
    end if
  end for

  //Updating
  for all Agent ∈ Population do
    r ← randomDouble()
    if r < pUpdate then
      Agent.updateLexicon()
    end if
  end for
end for
```

3.3 Updating

When selecting which of the incoming lexicons to incorporate into their lexicon, agents can use either *random* or *elitist* strategies [32]. The random strategy selects between each incoming convention seed uniformly at random, whereas the elitist strategy picks the seed with the highest quality. Salazar et al. [30] assume that agents send a quality valuation with the convention seeds, and that this is honest. In our investigation we adopt this assumption, but note that this is idealistic. The quality valuation for a lexicon (whether partial or complete) is the sum of the communicative efficacy and specificity for the full lexicon. As such, the two components are evenly weighted.

An agent's strategy is encapsulated by the population-wide variables p_{send} and p_{update}, the agent-specific variable l, its individual lexicon, and the agent-specific update strategy of random or elitist.

IAs are modelled as agents with a fixed lexicon: they attempt to propagate their own lexicon as normal (which may or may not be shared with other IAs), but will discard any incoming partial lexicons, regardless of their quality. We consider proportions of 0.05 IAs in a population (e.g. 50 agents out of 1000) to be an approximate of the upper bound on how many agents it might be practical to insert into a system in real-world domains, but for evaluation purposes we performed simulations with proportions up to 0.4. While such

proportions are likely to be impractical, it is useful to characterise the behaviour of the model at these settings.

3.4 Network topology

Agents are situated on a connecting topology which constrains communications to the immediate neighbour set. We use the Java Universal Network/Graph library (version 2.0.1)[4] to generate connecting topologies.

Our scale-free topologies are generated using the Eppstein and Wang power-law generator based on Eppstein and Wang's algorithm [9]. This differs from Barabási and Albert's [5] model of incremental growth, by instead evolving a graph with constant size and density using a Markov process. The algorithm takes three parameters, the total number of vertices n, the total number of edges e, and the number of edge insertions/deletions r. We use a value of $r = 1000000$.

Small-world topologies are generated using the Kleinberg small-world generator, based on Kleinberg's model [19]. In this model, an $m \times n$ lattice is augmented with a number of extra connections chosen with probability $p \propto d^{-\alpha}$, where α is the clustering exponent (CE), a parameter to the model, and d is the lattice distance between the two nodes being considered for a new edge. Our implementation here differs from Salazar et al. [30], who generated small world topologies using the Watts-Strogatz beta model and scale-free topologies using the Barabási-Albert model. More recent versions of JUNG do not include a generator for the Watts-Strogatz model. There are significant structural differences in networks generated by each model, most notable of which is that the Kleinberg generator tends to produce significantly lower clustering coefficients and networks with low total edge numbers. The two main parameters to this algorithm are the lattice size, and CE, and we use the default configuration in which each node is augmented with one extra connection. Unless otherwise stated, we use a 10×100 lattice with a CE of $\alpha = 5$.

3.5 Metrics

There are a number of important metrics which help characterise the efficacy and efficiency of convention emergence:

1. *Communicative Efficacy:* The average communicative efficacy of the system at each generation measures the ability of agents to communicate with each other effectively, and thus acts as a proxy for the level of coordination within the system.
2. *Number of agents using the most common lexicon:* As our results show, agents rarely all agree upon a single lexicon. The number of agents sharing the most commonly used lexicon indicates the level of convention adherence in the population.
3. *Distance of the most common lexicon from the initial IA lexicons:* The distance between two lexicons is defined as the number of mappings which are different. We are interested in the ability of IAs to influence the convention that is adopted throughout the entire population. If the most commonly used lexicon in the population is that used by the IAs, then we can consider the IAs to have been successful.
4. *Number of groups of a given size:* We define a group as a set of agents sharing an identical lexicon (and thus adhering to the same convention). Considering the number of groups and their sizes allows us to explore the evolution of conventions and sub-conventions over time. For example, if there are 1000 groups of size 1, then each agent has its own

[4] http://jung.sourceforge.net/.

unique lexicon and no conventions have emerged. One group of 750 agents and many other groups of small size would indicate one dominant, commonly-used convention, but with the rest of the population fragmented without strong convention emergence.

4 Results and discussion

4.1 Convention emergence without influencer agents

In order to determine the baseline system dynamics we initially consider the behaviour of the model without IAs. Unless otherwise stated, the results in this section are from running the simulation for 100000 timesteps using varying combinations of elitist and random lexicon selection strategies, and on a variety of topologies that reflect features of real-world domains. As discussed above, this represents a basic implementation of the model used by Salazar et al. [30] in which agents do not have the additional components of innovation, self-protection or noise, and our results corroborate the general conclusions of their work. Since we consider copy transfer to be impractical, all results reported here use partial transfer of lexicons. All results are averaged over 30 runs, except where single runs are used for illustration, using the parameter values specified in Table 1.

Figure 2a shows the average communicative efficacy over time for 1000 agents situated on a scale-free network topology with 10000 edges, while varying the proportion of elitist strategies in the population. For clarity, only the even proportions of elitist strategies have been plotted with symbols. The other (odd proportion) data sets are plotted as dashed lines without symbols. As the proportion of elitist agents increases, we see gains in the rate of increase of communicative efficacy and the final communicative efficacy the system converges to. All agents not using elitist strategies use random lexicon selection. Note that with 10000 edges, the graph is rarely fully connected and thus the probability of 100% adherence to the same convention is negligible, since the lexicon would have to emerge independently in each component. We further use low probabilities for lexicon spreading and update, and nodes with low degree are likely to receive far fewer partial lexicons with which to update. Consequently, these low-degree nodes may not have had sufficient chance to conform to the dominant lexicon by $t = 100000$. A population of entirely random selection agents (Elitist $= 0.0$) performs extremely poorly, with an average communicative efficacy remaining around 0.3. Increasing the proportion of elitist strategies through 0.1, 0.2 and 0.3 (the first three data sets above 0.0) shows significant gains both in terms of speed of increase of communicative efficacy and the final level reached after 100000 time steps. Further increasing the proportion of elitist strategies up to 0.9 results in further significant gains in the rate of increase of communicative efficacy, but there is little increase in the final value attained, and the final value the system converges to appears to be limited to an average communicative

Table 1 Default parameters used for the simulation configuration

Parameter	Value
Number of mappings in lexicon	10
P_{send}	0.001
P_{update}	0.001
t (timesteps)	100000
n (population size)	1000

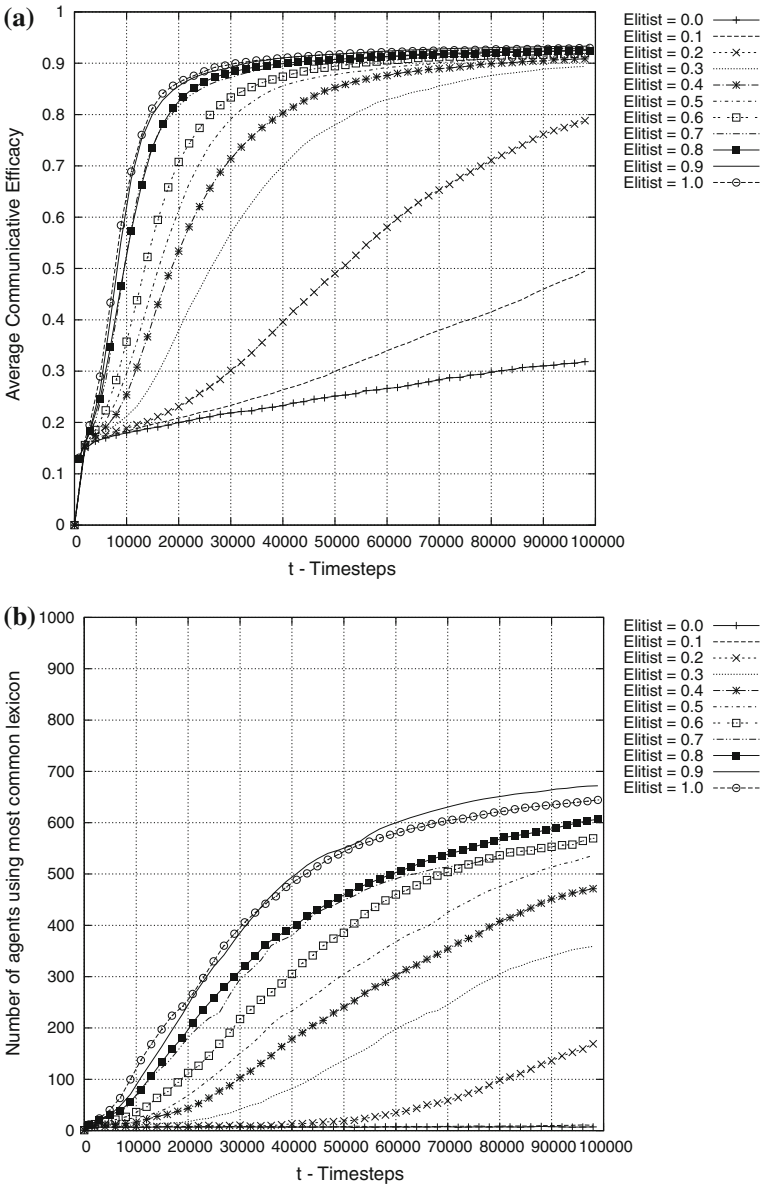

Fig. 2 Results showing **a** average communicative efficacy, and **b** the number of agents using the most common lexicon using a scale-free topology with 1000 agents and 10000 edges, with varying proportions of elitist and random lexicon update strategies in the population

efficacy of around 0.9. Applying a two tailed t-test with $\alpha = 0.05$ to the data shows that the difference in communicative efficacy between proportions of 0.8 and 0.9 elitist agents is significant with p-value 0.0316. Interestingly, the system does not necessarily perform better when all agents use an elitist strategy than when a 0.9 proportion of elitist strategies are present, with the results not being significantly different, with a p-value of 0.7697.

Figure 2b illustrates this more clearly and shows the number of agents using the most common lexicon in the population (i.e. the *dominant* lexicon) over time. Using an elitist proportion of 0.9 results in a larger group of agents using the dominant lexicon than with fully elitist populations. Overall, the differences in the number of agents using the most common lexicon are not statistically significant between elitist proportions of 0.9 and 1.0, with a p-value of 0.3528. The two sets of runs diverge at around $t = 50000$ time steps, but the difference is small, with a t-test for data at $t = 99900$ (i.e. an arbitrary time step after divergence) giving a p-value of 0.294. We hypothesise that the effectiveness of a 0.9 elitist proportion is due to the random selection agents introducing an element of exploration of lexicons into the system. The quality of a lexicon is given for the lexicon as a whole, but it may still be beneficial to allow parts of lower quality lexicons to spread, in a similar manner to mutation being beneficial in genetic algorithms. These partial lexicons may be helpful when included in other lexicons but may not be selected by elitist agents due to their inclusion in a lexicon of overall lower quality. [30] reported increased system performance when incorporating notions of controlled noise into agents' internal convention generation mechanisms, adding weight to this hypothesis. However, this effect may also be exacerbated by a limitation of purely elitist populations: when lexicons have many identical mappings, it takes more communications to determine a difference, thus slowing convergence. The small population of random selection agents thus reduces stagnation in populations containing similar lexicons.

4.2 The effect of network topology without influencer agents

We ran our simulations on small-world and scale-free networks with a variety of parameters. Figure 3a shows the effects of varying the total number of edges on a scale-free topology[5] from 1000 (very low) to 100000 (very high) with a population of entirely random selection strategy agents. We can see that at very low values for the total number of edges (1000) the population quickly converges to a stable average communicative efficacy of just under 0.4, despite using random selection strategies. At such low edge numbers, the network graph is highly disconnected. Increasing the number of edges significantly reduces this rate of convergence, since a higher number of edges results in a larger number of neighbours for each agent. With larger numbers of neighbours, each agent communicates, receives and sends to a larger group of other individuals, and the number of individuals with whom the agent has to agree on a convention for effective communication is increased.

Figure 3b shows results using the same parameters as in Fig. 3a, but with fully elitist populations. We again see that the number of edges strongly influences both the speed of convergence and the final communicative efficacy reached, with agents in a network with 100000 edges reaching a perfect lexicon. The fact that a perfect lexicon is reached is significant, since this corroborates results presented by [32], and also indicates the significant role that the underlying communication topology plays in convention emergence. There are a number of factors that are influenced by the number of edges in a scale-free topology, such as average node degree and average shortest path length. Given a larger number of edges, agents will not only have larger local neighbourhoods, but will also have shorter average path lengths to a larger cross-section of the society. Both factors allow high-quality conventions to be spread to a larger sub-set of the population more efficiently.

Figure 4 shows the average communicative efficacy for populations situated on a small world topology while varying the CE. The results are shown for populations of 100% Random

[5] The total number of edges in a scale-free network is the main parameter for the Eppstein and Wang generating algorithm we use [9].

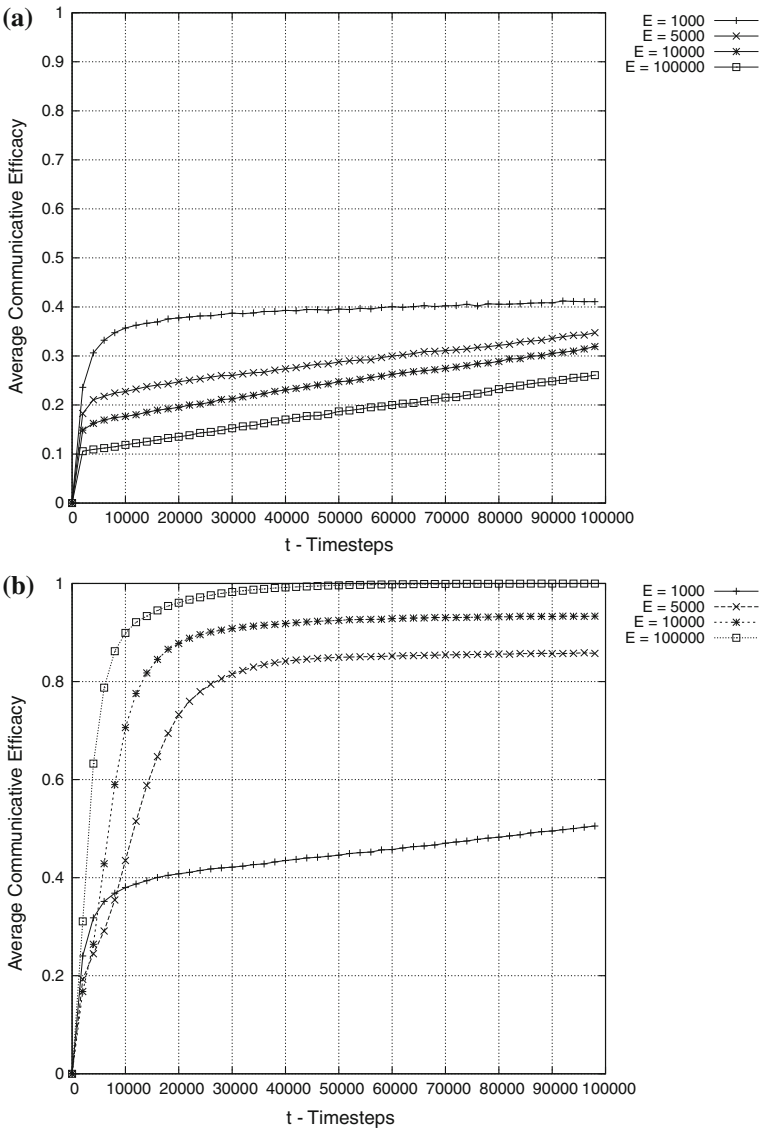

Fig. 3 Average communicative efficacy for a population of 1000 **a** random selection and **b** elitist selection strategy agents while varying the total number of edges in a scale-free topology

selection and 100% Elitist selection, and we can clearly see that (i) elitist populations deliver significant benefits over random selection populations, as with scale-free networks, (ii) the CE has a negligible effect on the dynamics of the simulation, and (iii) convergence is much slower on small-world networks than on scale-free networks (shown in Fig. 3), as reported by Salazar et al. [32]. It may be that the faster convergence on scale-free networks is due to the presence of hub nodes, which act to connect disparate clusters across the network, and thus may allow convention seeds to spread more effectively than on small-world networks that lack these features.

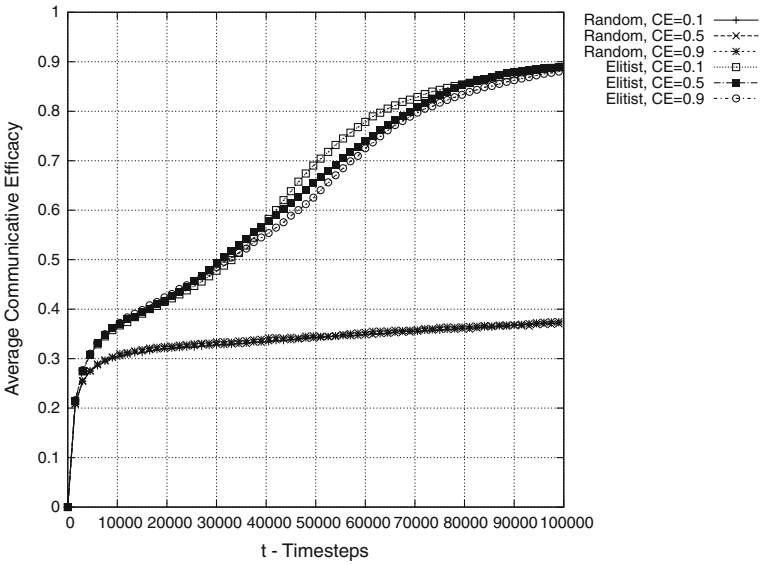

Fig. 4 Average communicative efficacy for 100% elitist and 100% random lexicon selection populations on small-world networks while varying the clustering exponent (CE)

Figure 5a and b plot the number of agents adhering to the dominant convention. Figure 5a shows the results for 100% Random and 100% Elitist populations on a small-world network while varying the CE. We can see that the small-world networks converge slower than comparable scale-free networks, and that both networks exhibit similar trends. We note that Salazar et al. [30] reported 100% adherence to the dominant convention for small-world networks. Our results are significantly different, in that despite corroborating the slower convergence of small-world networks we witness highly fragmented populations, with only around 100 to 150 agents adhering to the dominant convention after 100000 timesteps. As noted in Sect. 3, we use the default configuration for the Kleinberg small-world network generator, which tends to generate networks with a relatively low number of edges. When we compare the performance of the model on small-world networks to scale-free networks with comparable numbers of edges we see similar behaviour, as illustrated in Fig. 5b, which shows results for a small-world network and three scale-free networks with comparable edge numbers. Note that due to the constrained y-axis, the plot for 5000 edges is only partially shown. The behaviour of this run is similar to other runs with higher numbers of edges, and we show this data for comparison only. Given that these runs have not converged after 100000 timesteps, we performed a number of simulations to determine whether these populations converge at all. Using varying populations of random and elitist strategy agents on scale-free and small-world networks with up to 500000 timesteps, the results show that populations with up to a 0.2 proportion of elitist agents have still not converged by $t = 500000$, but that elitist proportions of 0.2 upwards have all converged by this time. As the proportion of elitist agents increases, the upper bound the population converges to and the speed of convergence both increase.

In order to analyse the evolution of the population in more depth, we can look at how *groups* of agents form over time. We define a group in this context as a set of agents adhering to the same shared lexicon. Initially, therefore, we expect there to be approximately 1000

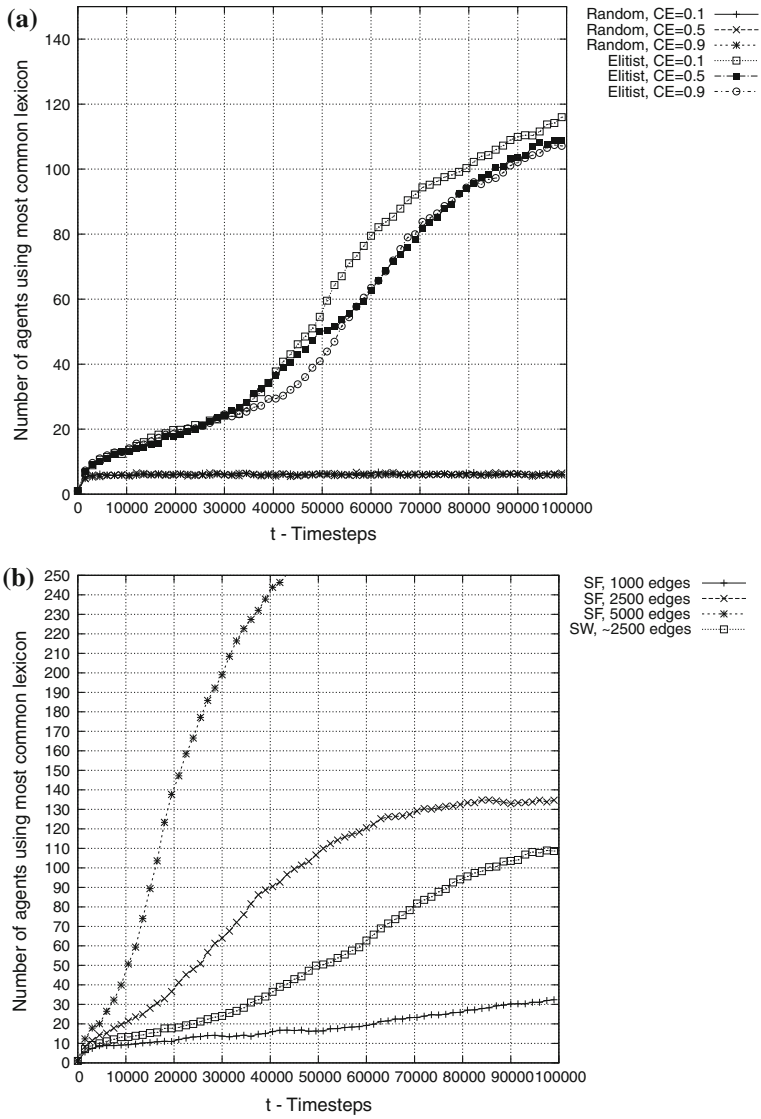

Fig. 5 Results showing the number of agents using the most common lexicon for **a** 100% elitist and 100% random lexicon selection populations on small-world networks while varying the clustering exponent (CE) and **b** comparable scale-free and small-world networks

groups, each containing a single agent. Our ideal aim in the language domain is for the population to achieve 1 group of 1000 agents, i.e. full convention emergence.[6] Table 2 details five snapshots of the system at $t = 0, 1000, 10000, 50000, 100000$, using fully elitist populations on a scale-free network with 10000 edges, for a single representative run. Figure 6 shows plots of the four later snapshots, but due to the difficulty of clearly plotting the whole

[6] As noted in Sect. 2, in some domains we might prefer a set of coordinated but distinct conventions, but consideration of such domains is outside the scope of this paper.

Table 2 Number of groups (n) of size (s), and average communicative efficacy (ACE), at various timesteps (presented as $n \times s$), for a fully elitist population of 1000 agents situated on a scale-free topology with 10000 edges

Timestep	Number of groups of size	ACE
0	1000×1	0.0
1000	922×1, 25×2, 2×4, 4×5	0.135
10000	298×1, 28×2, 11×3, 5×4, 1×5, 1×6, 1×8, 2×9, 1×10, 1×12, 1×13, 1×14, 1×17, 1×18, 1×19, 1×23, 1×28, 1×31, 1×37, 1×40, 1×47, 1×66, 1×77, 1×104	0.706
50000	34×1, 1×2, 1×22, 1×23, 2×25, 1×27, 1×29, 1×31, 1×32, 1×106, 1×644	0.757
100000	33×1, 1×20, 1×22, 4×23, 1×28, 1×29, 1×41, 1×735	0.934

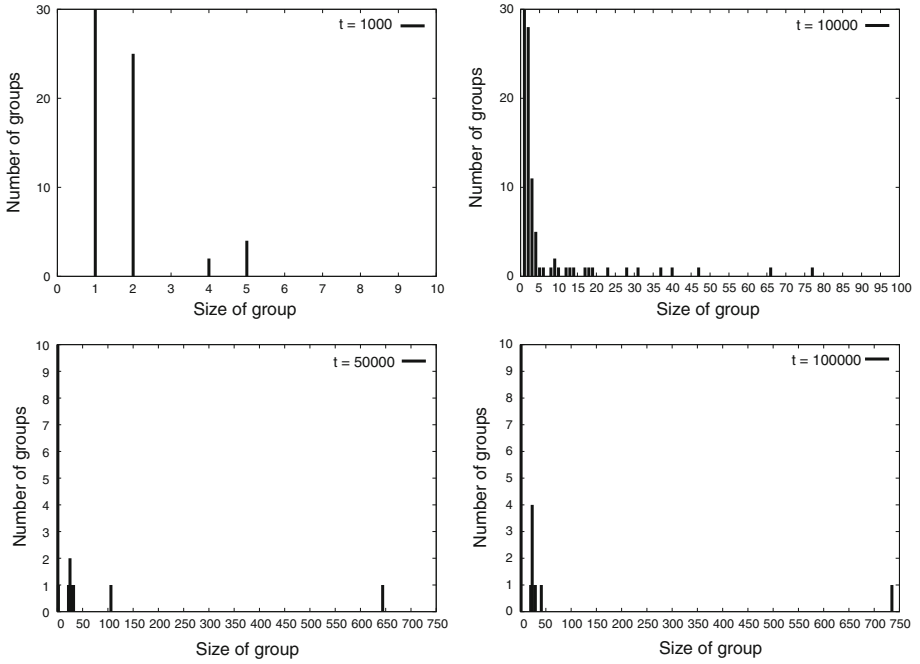

Fig. 6 Distribution of agent lexicon groups at $t = 1000, 10000, 50000$, and 100000, using a fully elitist population of 1000 agents situated on a scale-free topology with 10000 edges. Axes change scale between graphs for clarity of illustration. Figures are ordered in time from *left* to *right* and *top* to *bottom*

data set these figures are for illustrative purposes and only show a subset of the data. The x-axis corresponds to group size and the y-axis shows the number of groups of that size that exist at that time. We can see that at $t = 0$, every agent belongs to its own group of size 1, indicating that no two agents start with the same lexicon. At $t = 1000$, some evolution is seen, with a large number of subconventions emerging, each having a small number of adopters. At $t = 50000$ and $t = 100000$, we can see the dominant evolutionary pattern for agent groups: one large group starts to form and grows steadily throughout the simulation, while the rest of the population remains highly fragmented with no significant competition

for the main group, aside from one group of just over 100 agents at $t = 50000$ which has dissolved by $t = 100000$. Our ideal goal of a single group is not fulfilled, with the dominant group converging to 735 members.

Figure 7 contains four screen-captures of visualisations of the evolution of groups during a typical run on a scale-free topology with 10000 edges. Each image shows only the agents using the lexicon that eventually becomes the most dominant (dark), or those that are directly connected to them (light). Edges are coloured dark if the agents being connected by that edge are both using the dominant lexicon. Note that since the set of agents displayed in each image is different, the layout of the network is different between images and we therefore cannot use these images to infer results based on node position. While the number of agents in the population, and edges between them, means that the images are cluttered, we can make the following observations by visualising the evolution of the groups in this way:

1. At the start of the simulation, agents using the same lexicon are rarely directly connected, but are rarely more than 1 hop away from each other. This may be a quirk of topological structure, but it is more likely that the nature of lexicon spreading and update with partial lexicons means that lexicons will move closer to each other in small topological areas, generating a correlation between topological distance and lexical distance. Initially lexicons are unlikely to have any mappings in common due to the large convention space. Since spreading of lexicons is solely between neighbours, agents at opposite ends of the network are therefore unlikely to be exposed to the same convention unless all agents in between also move towards this convention.
2. Lexicons that go on to become dominant usually gain one or more high-degree adherents early on in the simulation, such that a large sub-set of the population gets exposed to the lexicon early on.
3. The nodes that are not part of the dominant lexicon at the end of the simulation tend to be low degree nodes on the fringe of the network.
4. The set of agents using a lexicon is rarely constant, and agents join and leave the lexicon convention constantly in the early stages. The lexicon that eventually dominates is distinguished therefore by having more agents join it each timestep than leave it.
5. The lexicon that eventually becomes dominant already exists in the population at $t = 250$. Although the use of partial transfer makes this seem surprising, we hypothesise that this is due to a combination of (i) low update and spreading rates and (ii) that an agent a using a lexicon successfully implies that *other* agents will update using a's seeds, spreading a's lexicon throughout the population (even though a might subsequently alter its own lexicon).

Analysis of the data regarding group formation and the topological illustrations suggests that the behaviour of the model on scale-free and small-world networks is markedly different. Scale-free networks are characterised by a single dominant group and the existence of fragmented sets of agents that do not adhere to the dominant convention. These agents typically exhibit the low node degrees that characterise locations at the fringes of the network. Conversely, in small-world networks the population tends to split into a few smaller groups, each one achieving high average levels of internal coordination. We have included the group data for small-world networks in the first column of Table 5, and therefore do not reproduce it here.

Convention emergence is faster and of higher quality in populations of elitist agents. Given that choosing the best convention is a rational decision, it is safe to assume that populations can be modelled as fully elitist. As such, in the remainder of this paper we use an elitist proportion of 1.0. The results presented above suggest that an elitist proportion of 0.9

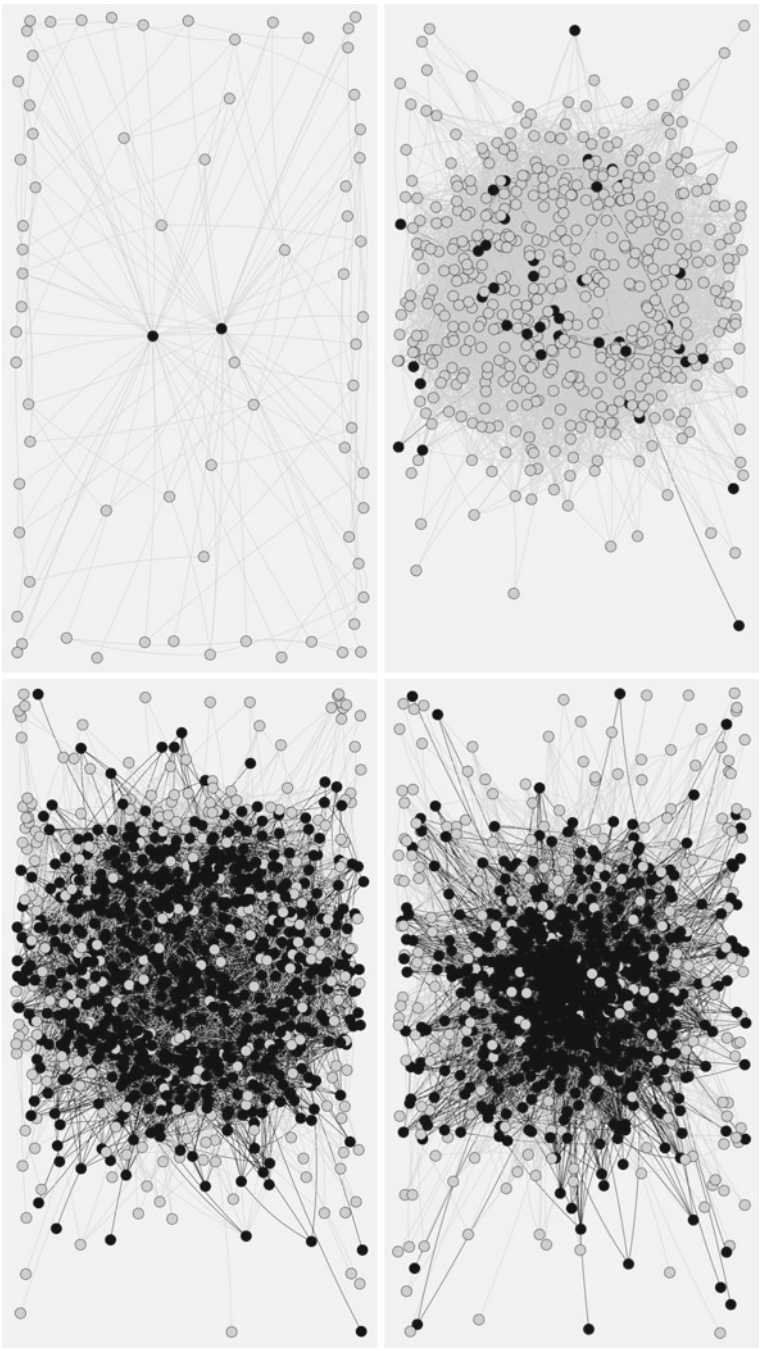

Fig. 7 Visualisation of group growth of the most dominant lexicon (by the end of the simulation) for a typical run at $t = 250$, 10000, 50000 and 100000 on a scale-free network. Agents using the lexicon in question are plotted dark, and their direct neighbours are plotted light. No other agents are plotted. Edges between two directly connected agents using the dominant lexicon are plotted dark. Figures are ordered in time from *left* to *right* and *top* to *bottom*

performs at least as well, but this would run contrary to our assumption of agent rationality and therefore is outside the scope of this investigation.

Our results in this section corroborate those presented by Salazar et al. [30], showing populations that converge to a high quality convention when using elitist lexicon update strategies. Small-world networks show slower rates of convergence than scale-free, although the convergence speed on the small-world networks presented here is significantly slower than that presented by Salazar et al. [30]. While these results validate our model and implementation with respect to Salazar et al. [30], we use different network topology generation algorithms. Our results for scale-free networks follow the same trends, implying that both the Barabasi-Albert and Eppstein and Wang generation algorithms generate networks with similar topological features. However, the behaviour we observe on small-world networks is markedly different, with populations fragmenting into multiple convention groups and no single dominant convention emerging. We believe this to be due to significant structural differences between the topologies generated by the Watts-Strogatz model and the Kleinberg model. Additionally, our analysis of system evolution by considering groups of agents adhering to single lexicons, shows that the typical route by which a population converges on scale-free networks is for a single lexicon to gain a large number of adherents quickly, but for the rest of the population to remain fragmented over many different groups (as opposed to adhering to a single competing lexicon, for example).

4.3 Introducing IAs

Now that we have established the baseline behaviour of the system, we introduce a small proportion of IAs to determine (i) if IAs can influence the dominant convention established in the population to that proposed by the IAs and (ii) if IAs provide any other benefits, in terms of speed of convergence and quality of convention reached, beyond being able to manipulate which convention the population agrees upon. To simplify analysis at this stage of the investigation, we assume that IAs are given a high quality (i.e. 1.0 specificity) lexicon at the start and that all IAs in the population share the same lexicon. IAs are randomly placed in the network topology. As discussed later in Sect. 4.5, the location of IAs in the social network does affect their ability to influence the population, but in this initial evaluation we are concerned with confirming their feasibility in general.

Figure 8a shows average communicative efficacy over time for various proportions of IAs, with 1000 agents on a scale-free network with 10000 edges. Figure 8b shows the number of agents using the most common lexicon in the population for the same set of runs. For small numbers of IAs, with proportions up to 0.005 (i.e. 5 IAs in our population of 1000), we can clearly see a significant increase in the speed at which agents start adhering to the most commonly used lexicon, and a marked increase in the rate at which communicative efficacy increases, although it converges to a similar value as without IAs. As such, it may be that the increased speed is merely due to the presence of a high quality lexicon in the population at the start of the simulation, and we discuss this possibility further below. At proportions of 0.1 (i.e. 100 IAs in a population of 1000) all runs end with the most commonly used lexicon being the lexicon given to IAs at the start (a result that holds for both scale-free and small-world topologies). However, such a high proportion of IAs is likely to be impractical in a real application domain. In Sect. 4.5, we show that targeting IA placement by node degree reduces the proportion of IAs necessary to guarantee convergence to the IA lexicon by a factor of 10, to just 0.01. When placed randomly, a proportion of 0.005 IAs results in 43% of runs ending with the most common lexicon being the initial IA lexicon. This supports our hypothesis that small groups of agents can significantly influence convention

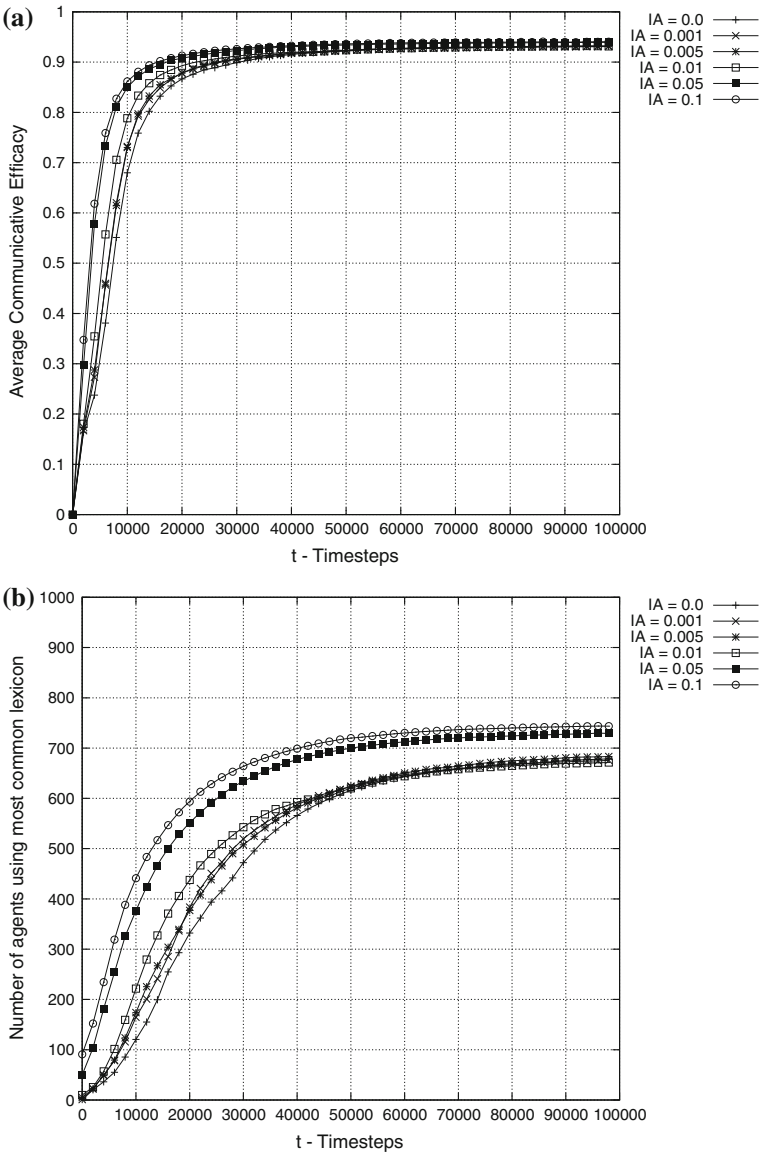

Fig. 8 Results showing **a** average communicative efficacy, and **b** the number of agents using the most common lexicon for varying proportions of IAs that are initialised with same high quality lexicon. Results are shown for a scale-free topology with 10000 edges. The non-IA population is entirely elitist

emergence in a large population. This finding corroborates that of [35] in which 4 agents could convert a population of 3000 to their convention, although the convention space they consider is only binary. We also find that 46% of runs (i.e. an additional 3%) end with the dominant lexicon having a distance of at most 2 from the initial IA lexicon (i.e. the lexicons differ by at most 2 mappings, indicating a close match), and that most of the runs which do not end in convergence to the IA lexicon have a distance of 9 or 10. These results are

highly polarised: either the IA lexicon is adopted (or in around 3% of cases, a very similar version adopted), or a completely different lexicon becomes dominant. This may be due to the initial group dynamics: if a large group of agents adopt a lexicon that is distant from the IA lexicon early in the simulation, then the influence of IAs over this group becomes insufficient. If the IA lexicon becomes dominant early on then the communicative efficacy is greatly increased, leading to convergence to the IA lexicon. Furthermore, the additional 3% of runs that end with a dominant lexicon very similar to the IA lexicon are likely to be a result of the partial transfer mechanism resulting in modified IA lexicons being propagated through the population.

We performed numerous *t*-tests to determine the significance of our results. Between IA proportions of 0.0 and 0.005 (i.e. between 0 and 5 IAs in 1000 agents), we find that the change in average communicative efficacy is not significant, with a *p*-value of 0.2564. However, the gain in the number of agents using the most common lexicon is significant, with *p*-value 0.03730. Between proportions of 0.0 and 0.01 (i.e. an addition of 10 IAs into a population of 1000), we find significance in both the gains in communicative efficacy and the number of agents using the most common lexicon, with *p*-values of 0.01254 and 0.00064 respectively. These results suggest that IAs can give significant benefits to a population beyond simply manipulating the convention that emerges.

Figure 9a and b shows the average communicative efficacy and the number of agents using the dominant lexicon respectively, for the same configuration as Fig. 8, but with very high proportions of IAs. As noted above, we consider these proportions to be impractical for real-world application, but they are useful for understanding the dynamics of the approach. There are further gains in the number of adherents to the dominant lexicon and the speed of convergence to the upper bound of lexicon adherents, and a very minor gain in the speed of convergence for average communicative efficacy, but the cost of these gains has markedly increased, given that each data set represents another additional 100 IAs inserted into the system.

Figure 10a and b shows the average communicative efficacy and the number of agents using the dominant lexicon, respectively, for the same configuration as Fig. 8, but on small-world networks. The data shows the same trends but with the typical slower convergence rates and lower upper bound on dominant lexicon adherence. We see similar gains in both the levels reached and the speed of convergence for both metrics. Our results imply that controlling convention emergence on small-world networks requires more IAs than on scale-free networks. As noted above, small-world networks appear to encourage convergence to multiple stable conventions, rather than a single dominant convention. Moreover, small-world networks are slower to converge than scale-free. The costs of controlling small-world networks are thus higher than scale-free, but these costs may be reducable through more refined IA strategies or targeting IA placement more effectively. Our results suggest that the presence of hubs, a key feature of scale-free networks, reduces convergence time.

Average communicative efficacy is a useful proxy for the level of coordination in a system, with higher levels of communicative efficacy indicating agents adhering to more similar conventions. The addition of IAs does not appear to increase the levels of average communicative efficacy attainable by the society, and as such we cannot say that we are increasing levels of coordination. However, we are not just interested in increasing levels of coordination with the IA mechanism. We are also interested in (i) determining the extent to which we can determine *which* convention a society adopts, (ii) increasing levels of adherence to the dominant convention, and (iii) understanding how to exploit the changes that small groups of agents can affect. Our results show that we can influence the convention that emerges in a society, and in doing so increase the levels of adherence.

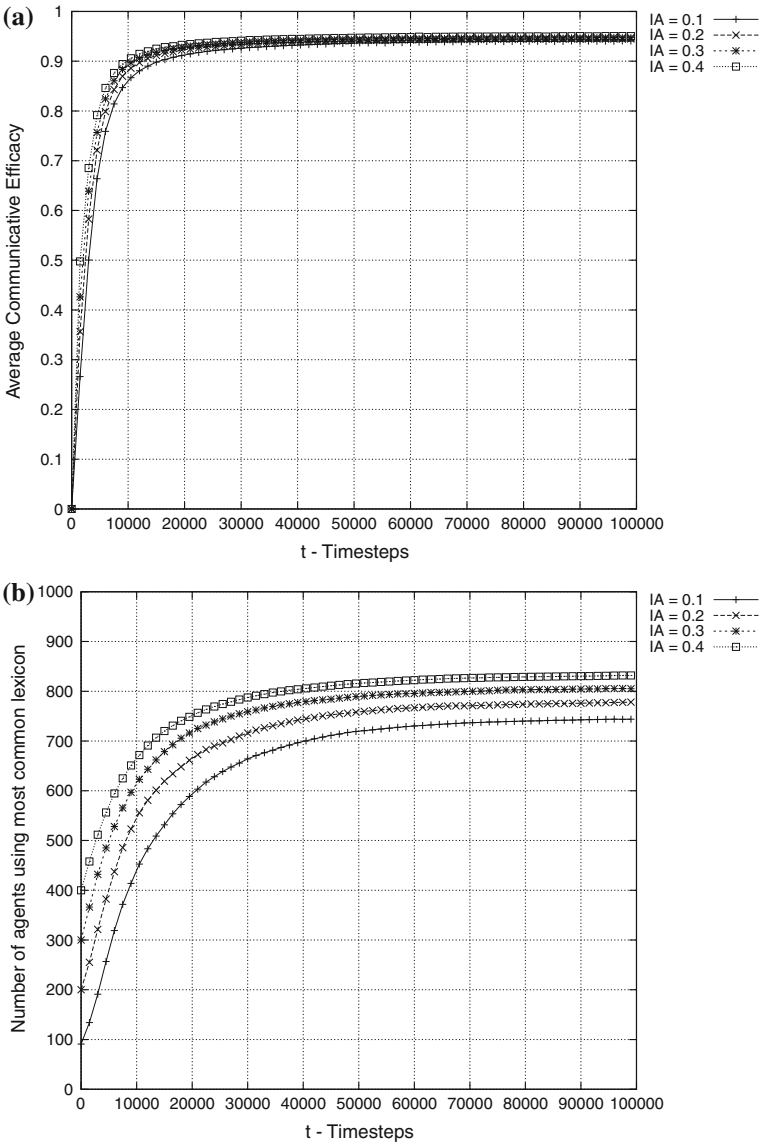

Fig. 9 Results showing **a** the average communicative efficacy and **b** the number of agents using the most common lexicon for very high proportions of IAs that are initialised with same high quality lexicon. Results are shown for a scale-free topology with 10000 edges. The non-IA population is entirely elitist

4.4 Inserting agents versus inserting conventions

The introduction of IAs into the population entails two major differences in the configuration of the model: (i) the existence of a small proportion of agents with a fixed strategy, and (ii) the existence of a high quality lexicon in the population at the start of the simulation. We are interested in manipulating convention emergence using the former, and thus we need to quantify the effects of the latter without the existence of fixed-strategy agents in the population.

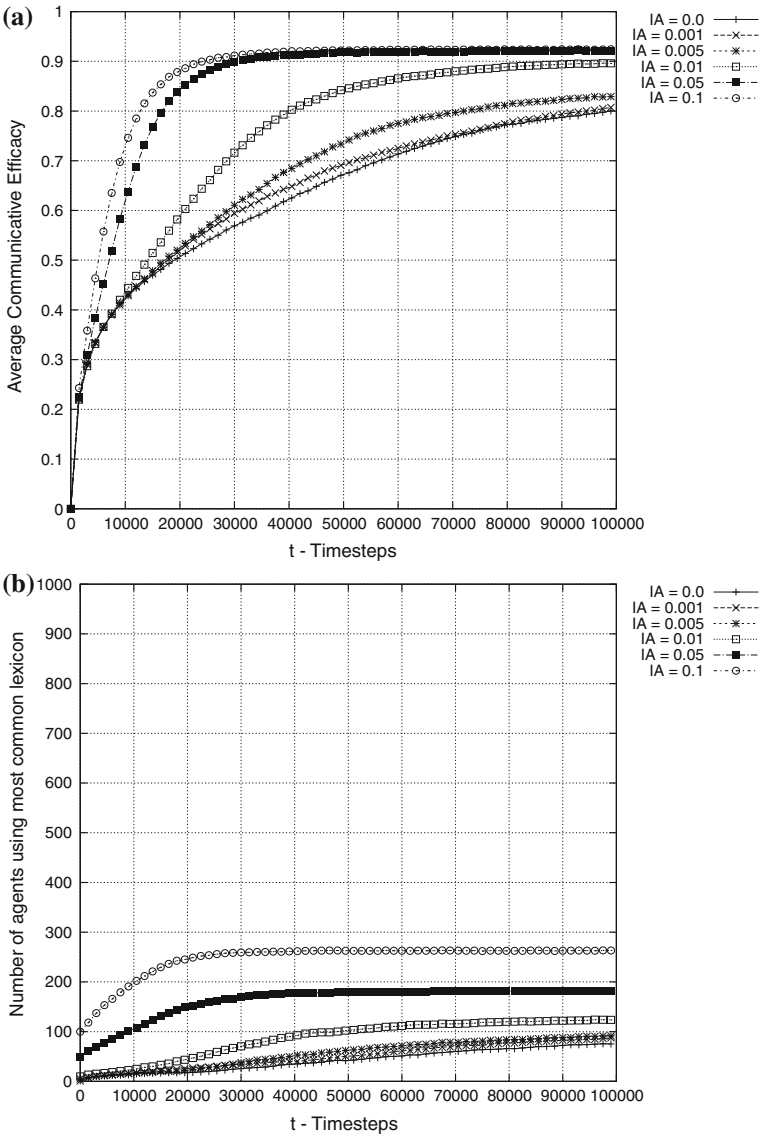

Fig. 10 Results showing **a** average communicative efficacy and **b** the number of agents using the most common lexicon for varying proportions of IAs that are initialised with same high quality lexicon. Results are shown for a small-world topology. The non-IA population is entirely elitist

We ran simulations in which we replace IAs with agents that are given the same high-quality initial lexicon, but in all other respects are identical to regular agents (i.e. they propagate and update their lexicon as normal). For clarity of discussion, we call these agents HQ (high-quality) agents. These simulations represent an implementation of the second model defined in Sect. 2.2, in that rather than using a small proportion of inflexible agents to continuously propagate a convention, we instead imbue a certain proportion of standard flexible agents with a high quality convention and determine how this affects convention emergence.

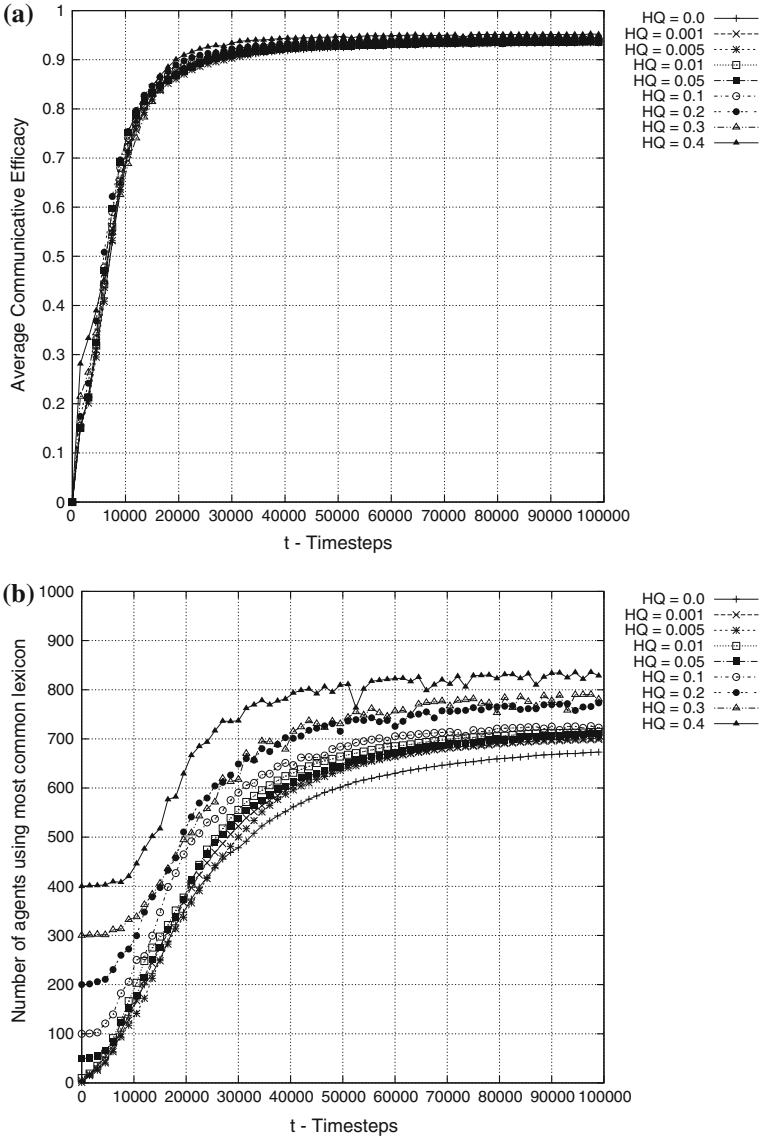

Fig. 11 Results showing **a** average communicative efficacy, and **b** the number of agents using the most common lexicon for varying proportions of agents initialised with the same high quality lexicon. Results are shown for a scale-free topology with 10000 edges

Figure 11a shows average communicative efficacy using varying proportions of HQ agents instead of IAs on a scale-free topology, and Fig. 11b shows the number of agents using the most common lexicon for the same set of runs. Comparing proportions directly with Fig. 8, we can see that at values between 0.01 and 0.1 (i.e. realistically insertable proportions), IAs and HQs show very similar behaviour. Comparing sets of runs with a HQ proportion of 0.005 and an IA proportion of 0.005, we see no significant difference in the number of agents using the most common lexicon (the data gives a p-value of 0.7798). However, when we compare

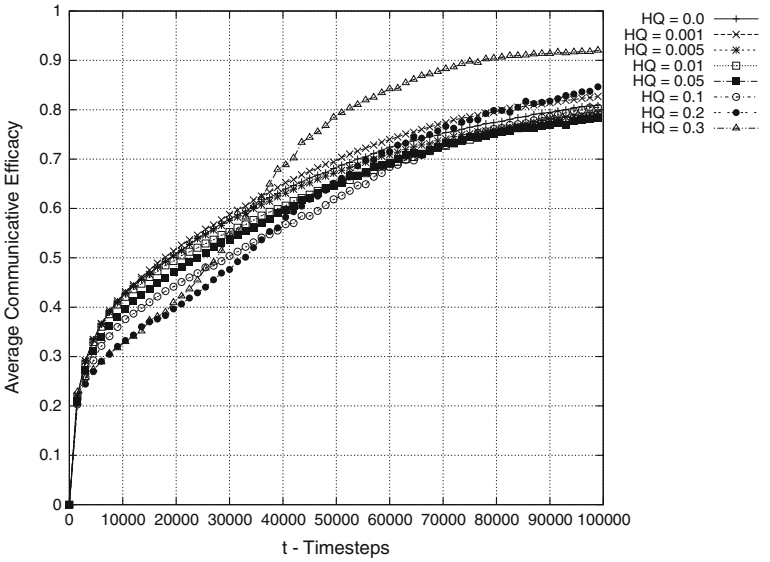

Fig. 12 Results showing the average communicative efficacy for varying proportions of agents given a high quality lexicon at the start of the simulation, situated on a small-world topology

10 HQ agents versus 10 IAs (i.e. a proportion of 0.01), we see a significant difference in the number of agents using the dominant lexicon, with a p-value of 0.01653, and HQ agents performing better than IAs with respect to the number of dominant lexicon adherents. As we increase the proportions, IAs show bigger gains in the number of adherents, with a t-test between runs with proportions of 0.1 (i.e. 100 HQ agents compared to 100 IAs in a population of 1000) giving a p-value of 2.642×10^{-16}. Despite this significance, the differences in the number of agents adhering to the dominant convention are never large (743 with IAs compared to 726 with HQ agents). Comparing the higher proportions of HQ agents with Fig. 9b, we can see that even at these proportions the differences are slight, though IAs give slightly better results. However, with HQ agents we *cannot control which convention emerges*, as we are simply injecting conventions into a population and letting the agents carry on as normal. Conversely, with IAs we can control the conventions that emerge, and also get the marginal gains in the metrics discussed above.

The gains that IAs show over HQ agents are similar with respect to communicative efficacy when comparing equal proportions, with the differences only becoming significant at a proportion of 0.05 (with a p-value of 1.057×10^{-5}). The gains in average communicative efficacy we see from introducing a high quality lexicon do not scale linearly with the proportion of agents given that lexicon. As Fig. 11a shows, we do not see any major improvements in average communicative efficacy on scale-free networks, even with very high proportions of HQ agents. This contrasts with the results we see in Fig. 12, in which increasing the proportion of HQ agents gives highly visible benefits on small-world networks. It is not clear whether the HQ agents increase the upper bound of average communicative efficacy that the society attains due to the slow convergence of the model on small-world networks, but nonetheless we can see significant increases in the speed of convergence on these networks. These results suggest that seeding small-world networks with high quality conventions initially might effectively reduce the additional costs associated with small-world networks discussed above. When

Table 3 Table showing whether a *t*-test of the levels of dominant convention adherence exhibited with different proportions of IA or HQ agents results in a statistically significant *p*-value ($\alpha = 0.05$), indicated by ∘, or not, indicated by •

HQ proportion	IA proportion						
	0.4	0.3	0.2	0.1	0.05	0.01	0.005
0.4	∘	∘	∘	∘	∘	∘	∘
0.3	∘	∘	∘	•	∘	∘	∘
0.2	∘	∘	∘	•	•	∘	∘
0.1	∘	∘	∘	∘	∘	∘	∘
0.05	∘	∘	∘	∘	∘	•	∘
0.01	∘	∘	∘	∘	∘	∘	∘
0.005	∘	∘	∘	∘	∘	•	•

Table 4 Statistical significance resulting from a *t*-test of the average communicative efficacy exhibited with different proportions of IA or HQ agents, with $\alpha = 0.05$, where • represents non-significant differences, and ∘significant differences

HQ proportion	IA proportion						
	0.4	0.3	0.2	0.1	0.05	0.01	0.005
0.4	∘	∘	∘	∘	∘	•	•
0.3	∘	∘	∘	∘	∘	•	•
0.2	∘	∘	∘	∘	∘	•	•
0.1	∘	∘	∘	∘	∘	•	•
0.05	∘	∘	∘	∘	∘	•	•
0.01	∘	∘	∘	∘	∘	•	•
0.005	∘	∘	∘	∘	∘	•	•

comparing the results for average communicative efficacy in Figs. 11a and 9a, we see that at very high proportions IAs result in much faster convergence than HQ agents.

Table 3 shows the results of statistical significance tests for the number of agents adhering to the most common lexicon between differing proportions of HQ and IA agents, placed randomly on a scale-free topology with 10000 edges. Where the *p*-value is significant, with $\alpha = 0.05$, we have marked the entry ∘, and where the *t*-test indicates that the data is statistically indistinguishable we have marked the entry •. It is interesting to note that 300 HQ agents (a proportion of 0.3) produces results statistically indistinguishable from 100 IAs (a proportion of 0.1), but that 5 HQ agents produce results statistically indistinguishable from both 10 and 5 IAs. It appears that at very low proportions the effects of HQ and IA agents are difficult to distinguish, with IAs increasing the levels of dominant convention adherence quicker than HQ agents. Table 4 shows results in the same format as Table 3, for statistical significance tests of the average communicative efficacy from the same set of runs. Here the data is clearer, and we can see that at low proportions IAs are indistinguishable in their effects from any proportion of HQ agents, as neither agent type produces major improvements, but that as we increase the proportion of IAs we see significant gains over HQ agents, although the actual improvement is fairly minor (see Fig. 11a).

Table 5 Number of groups (n) of size (s) at various timesteps (presented as $n \times s$) for representative runs with a 0.0, 0.005, and 0.01 proportion of IAs inserted by highest node degree on a small-world topology

Timestep	Number of groups of size		
	$IA = 0.0$	$IA = 0.005$	$IA = 0.01$
0	1000×1	$995 \times 1, 1 \times 5$	$990 \times 1, 1 \times 10$
1000	902×1	897×1	909×1
	$28 \times 2 \leq s \leq 4$	$39 \times 2 \leq s \leq 4$	$33 \times 2 \leq s \leq 4$
	2×5	2×5	1×12
10000	673×1	681×1	635×1
	$76 \times 1 < s < 5$	$69 \times 1 < s < 5$	$65 \times 1 < s < 5$
	$19 \times 5 \leq s < 15$	$21 \times 5 \leq s < 8$	$18 \times 5 \leq s < 15$
	1×22	2×20	$1 \times 15, 2 \times 17$
			1×25
50000	368×1	320×1	47×1
	$61 \times 1 < s < 5$	$61 \times 1 < s < 5$	$28 \times 1 < s < 15$
	$41 \times 5 \leq s < 15$	$14 \times 5 \leq s < 15$	$2 \times 15 \leq s < 50$
	$7 \times 15 \leq s < 25$	$16 \times 15 \leq s < 35$	$8 \times 50 \leq s < 100$
	1×27	1×42	$1 \times 100, 1 \times 107$
100000	171×1	6×1	2×1
	$28 \times 1 < s < 5$	$5 \times 1 < s < 5$	$5 \times 1 < s < 5$
	$54 \times 5 \leq s < 25$	$14 \times 5 \leq s < 50,$	$16 \times 5 \leq s < 50$
	$10 \times 25 \leq s < 45$	$7 \times 50 \leq s < 85,$	$7 \times 50 \leq s < 100$
	1×46	1×89	$1 \times 100, 1 \times 114, 1 \times 115$

4.5 Effect of position of IAs in the network

The experiments detailed above involve IAs placed in random locations in the social network structure. While this may be all that is possible in some domains, an obvious modification involves targeting the location of IAs to improve their efficacy. Intuitively, some locations in a network are more influential than others, and placing IAs at more influential locations should improve their efficacy.

To confirm whether topological properties might influence the effectiveness of IAs, we ran simulations with varying IA proportions for 1000 agents on scale-free and small-world networks, in which IAs are (for simplicity) given a high quality lexicon initially. Instead of randomly placing IAs, we place them according to node degree, a commonly used approximation for influential locations in social network analysis (for example see [7,17]).

We found that when placing a 0.005 proportion of IAs at locations with the highest node degree in a scale-free topology, 66% of runs ended with the most commonly used lexicon being at most a distance of 2 from the initial IA lexicon. The average distance over 30 runs is 2. This compares favorably with results reported earlier in Sect. 4 in which 46% of runs with randomly placed IAs ended with the most commonly used lexicon being at most a distance of 2 from the initial IA lexicon, and confirms the hypothesis that topological properties significantly affect the efficacy of IAs. When we place IAs at locations in the topology with the lowest node degree, we find that only 20% of runs end with the most commonly used

lexicon being at most a distance of 2 from the IA lexicon. Furthermore, the average distance over 30 runs is 7.

On both scale-free and small-world topologies we find that a 0.01 proportion of IAs (i.e. 10 agents in 1000), when placed by node degree, is enough to result in 100% of runs ending with the most commonly used lexicon being at most a distance of 2 from the initial IA lexicon, demonstrating the power of topologically informed placement. Using knowledge of topological properties such as node degree, IAs may be able to incorporate re-wiring strategies to move themselves to more influential positions in the connecting topology of an artificial society, although such considerations are beyond the scope of this paper.

Figure 13a and b shows the average communicative efficacy and the number of agents adhering to the most common lexicon respectively for varying proportions of IAs when placed by node degree on a scale-free network. Figure 14a and b shows the corresponding results for a small-world network. As before, IAs exhibit diminishing returns at high proportions (i.e. 0.1 to 0.4), implying that the gains we see at low proportions are close to the upper bound. At these high proportions, placement by node degree does not lead to significantly different values for our metrics than random placement, and so we have not included figures for these data. It is interesting to note that the number of agents adhering to the dominant convention falls slightly when placing agents by node degree (as opposed to randomly), despite the ability of IAs to control which convention emerges increasing.

Table 5 shows group data for representative runs inserting various proportions of IAs by node degree in a population of 1000 agents on small-world networks. Note that since the group structures are markedly different between IA proportions, the group sizes between which we aggregate group numbers change between table cells. We can see that in small-world networks, as discussed above, agents tend to form small groups adhering to conventions, rather than emerging one single dominant convention. The insertion of 5 IAs has only a small effect, with a larger dominant group size and more groups at the larger group sizes. There appears to be a threshold between 5 and 10 IAs at which the IAs start having a more profound impact on group formation, with many more groups forming with higher numbers of adherents and the emergence of three groups with 100 or more agents. The data illustrates the slower emergence of conventions on small-world networks than on scale-free, but despite this IAs still have a visible and beneficial effect on the conventions that do emerge.

It is important to note that node degree is strongly related to observability of agents, in that a more highly connected agent is able to spread convention seeds to more agents. As such, we cannot trivially generalise these results to other topological properties such as the local clustering coefficient, which have more complex relationships with respect to the propagation of information through the network, and such consideration is left for future investigation.

4.6 IAs with imperfect conventions

Our evaluation of IAs has so far assumed, for simplicity, that they are given the best possible initial lexicon. In realistic domains, in which it may not be possible to identify ideal conventions a priori, we cannot assume that this is the case. In this section, we investigate the effects of introducing IAs with entirely randomly generated lexicons. As such, the quality of these lexicons is likely to be poor on average—in a sample of 200 randomly generated lexicons, the average specificity was 0.521, with standard deviation 0.129. Since we assume that the rest of the population is elitist, it is unlikely that these lexicons will be adopted, but it is important to explore whether the presence of agents with fixed poor strategies impedes the emergence of high quality conventions. In any real world setting, although we may not be able to generate a perfect convention initially, we do expect that IAs will be able to adapt

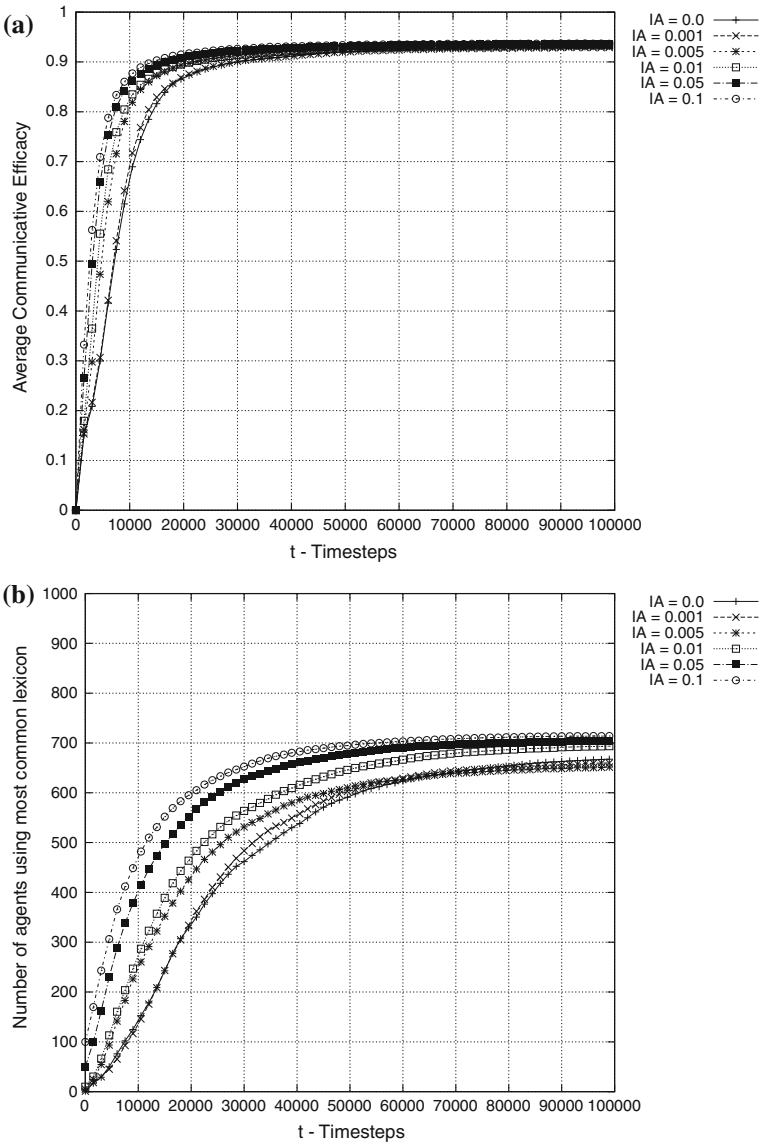

Fig. 13 Results showing **a** average communicative efficacy, and **b** the number of agents using the most common lexicon for varying proportions of IAs when placed in the topology by node degree. Results are shown for a scale-free topology with 10000 edges. The non-IA population is entirely elitist

themselves and thus still aid the emergence of high quality conventions. Such adaptations in response to changes in the environment are outside the scope of this paper.

Figure 15a shows the number of agents using the most common lexicon, on a scale-free topology, for IA proportions from 0.0 to 0.1. All IAs are given the *same* randomly generated lexicon. As discussed above, these lexicons have low average specificities. At the start of the simulations, there is little difference between runs, as the average quality of lexicons in the whole population is also poor. However, as the system progresses, we can see the runs start

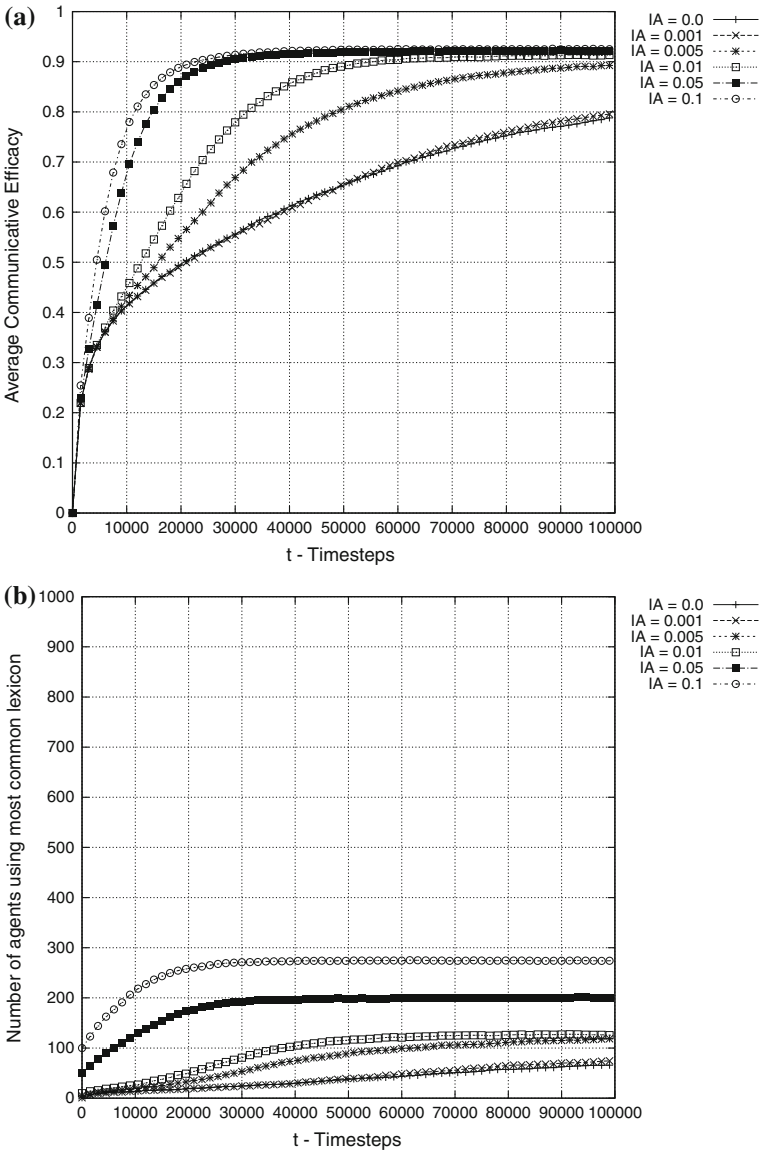

Fig. 14 Results showing **a** average communicative efficacy, and **b** the number of agents using the most common lexicon for varying proportions of IAs when placed in the topology by node degree. Results are shown for a small-world topology. The non-IA population is entirely elitist

to diverge, with configurations containing higher proportions of IAs exhibiting significantly fewer agents adopting the dominant lexicon. There is clearly some detrimental effect on the levels of convention emergence in the population, with fewer agents joining the dominant convention, and at a slower rate. It is interesting to note that a proportion of $IA = 0.001$ appears to perform slightly better than a proportion of $IA = 0.0$, but the difference is not statistically significant. Figure 15b shows results using the same configuration as Fig. 15a,

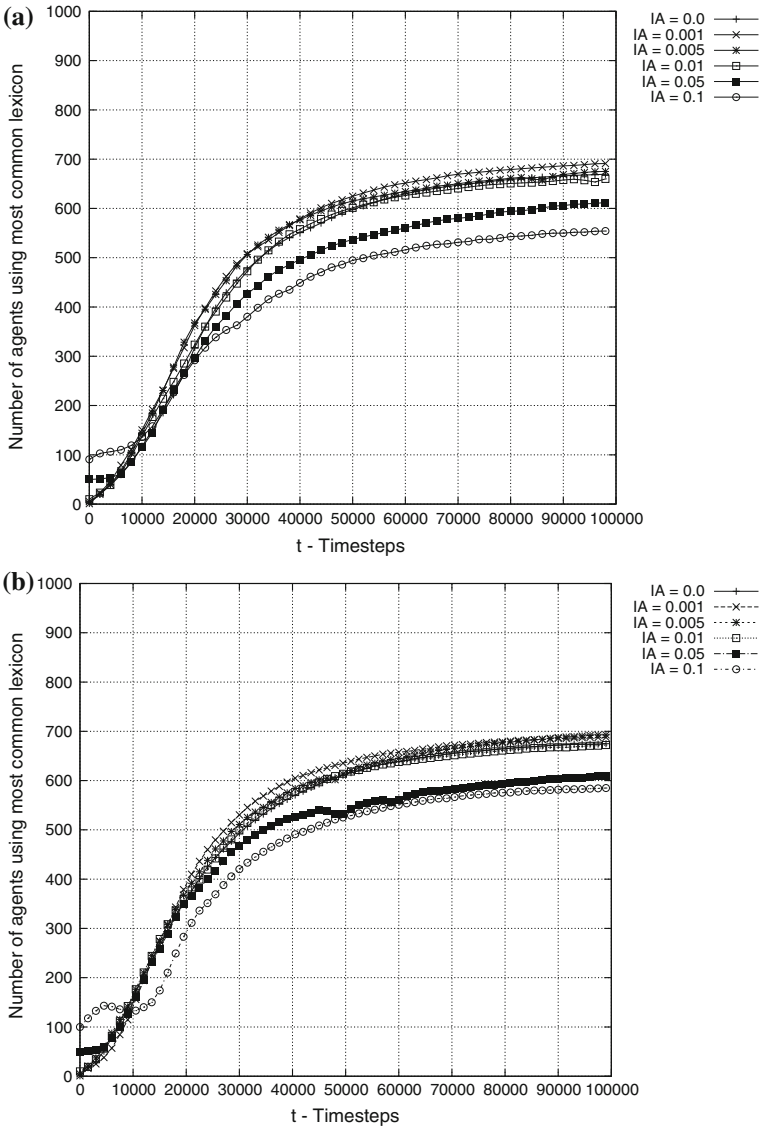

Fig. 15 Number of agents using most common lexicon on scale-free topology with 1000 agents and 10000 edges, while varying proportion of IAs. In **a** IAs are given the same random (and therefore, on average, poor quality) lexicon, whereas in **b** IAs are given different, unique random lexicons

but instead giving each IA a *different* randomly generated lexicon. We can see here similar detrimental effects, in that as we increase the numbers of IAs with poor quality lexicons we see a decrease in the number of adherents to the dominant convention, but the size of the decrease is lower when IAs propagate different poor quality lexicons. A small set of IAs propagating poor quality lexicons can be viewed as equivalent to low levels of noise in the model. When IAs have different poor quality lexicons, increasing the proportion of IAs increases the levels of noise, with corresponding detrimental effects. Conversely, when

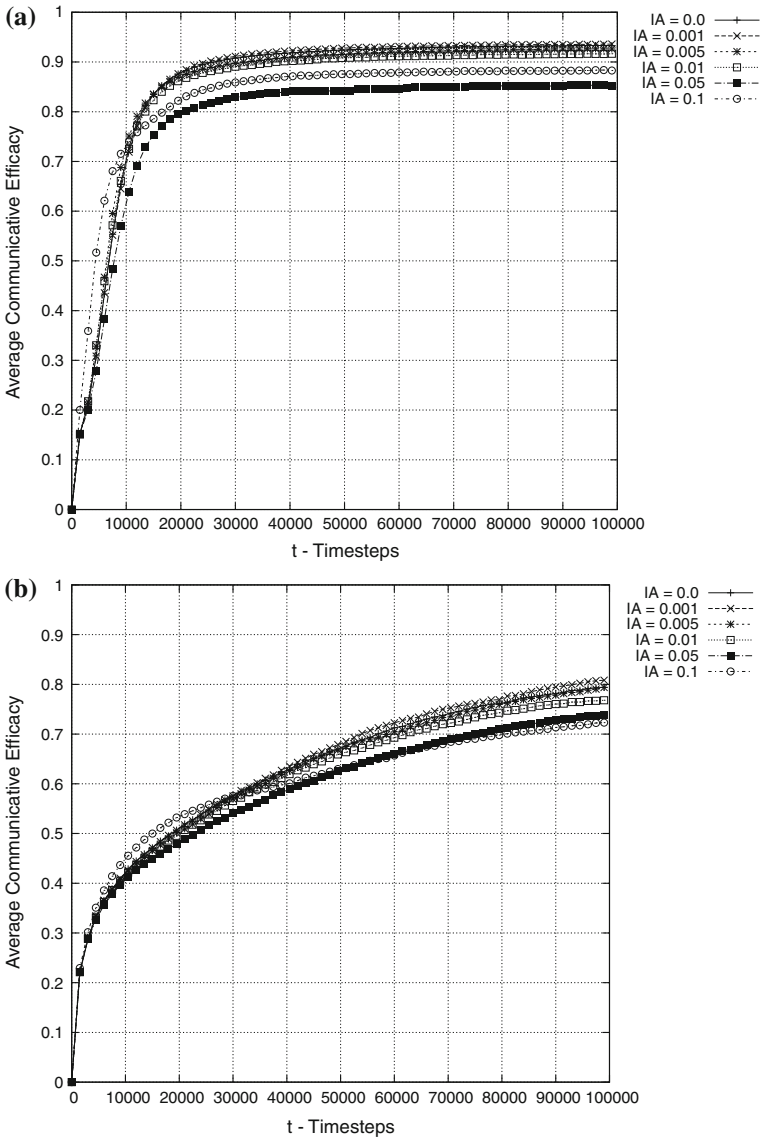

Fig. 16 Results showing the average communicative efficacy for **a** scale-free and **b** small-world networks while varying the proportion of IAs. IAs are given the same random lexicon

the IAs use a single poor lexicon, increasing the proportion of IAs constitutes a coordinated malicious effort by the IAs to disrupt convention emergence, and the detrimental effects are larger.

Figure 16a and b shows the average communicative efficacy for scale-free and small-world topologies respectively, with IAs given the same poor quality lexicon. We can see similar effects, as increasing the number of IAs with poor lexicons significantly reduces the level of average communicative efficacy. It is interesting to note that on scale-free topologies the

Table 6 Number of groups (n) of size (s) at various timesteps (presented as $n \times s$) for representative bad, average, and good runs, for an IA proportion of 0.05 inserted randomly on a scale-free topology with 10000 edges

Timestep	Number of groups of size		
	Bad run	Average run	Good run
0	950×1, 1×50	950×1, 1×50	950×1, 1×50
1000	886×1	864×1	862×1
	$22 \times 1 < s < 20$	$22 \times 1 < s < 20$	$30 \times 1 < s < 20$
	1×52	1×50	1×52
10000	309×1	331×1	625×1
	$70 \times 1 < s < 20$	$75 \times 1 < s < 20$	$72 \times 1 < s < 20$
	$5 \times 20 < s < 50$,	$8 \times 20 < s < 50$	$1 \times 20 < s < 50$
	1×50	1×51	1×51
	1×64	1×58	
50000	44×1	48×1	51×1
	$4 \times 1 < s < 20$	$4 \times 1 < s < 20$	$0 \times 1 < s < 20$
	$4 \times 20 < s < 50$,	$6 \times 20 < s < 50$	$9 \times 20 < s < 50$
	$6 \times 50 < s < 90$	$3 \times 50 < s < 90$	1×52
	1×346	1×481	1×593
100000	41×1	40×1	34×1
	$3 \times 1 < s < 20$	$2 \times 1 < s < 20$	$0 \times 1 < s < 20$
	$5 \times 20 < s < 50$,	$7 \times 20 < s < 50$,	$9 \times 20 < s < 50$
	$6 \times 50 < s < 90$,	$3 \times 50 < s < 90$,	1×51
	1×445	1×566	1×640

IAs are given the same random, and therefore poor quality, lexicon

communicative efficacy actually increases from IA proportions of 0.05 to 0.1. This may be due to the effect of the communicative efficacy between IAs themselves, being given the same lexicon, beginning to have a significant impact on the overall average.

Analysis of the data shows significant differences in system dynamics between individual runs, where before (in configurations discussed in the previous sections) there had been no significant differences between runs. Table 6 shows the group evolution in three representative runs which we have classified as good, bad or average according to the number of agents adopting the dominant convention by the end of the simulation. For clarity, groups have been aggregated between intervals instead of listing every individual value. We can see that while the fragmentation of the population into groups is roughly similar at the start of the simulation, as time progresses the bad run remains highly fragmented and with a much smaller dominant group than that attained in the good run.

There are only two differences between these individual runs, given that they all have the same parameter settings: (i) the set of lexicons given to agents at the beginning, and (ii) the connecting topology and location of IAs, which are entirely random. Clearly, one of these factors must account for the disparity in convention emergence between individual runs.

To understand the cause of these differences we begin by considering the effect of the initial lexicons given to agents. Simulations without IAs do not show such a significant disparity between individual runs, so the differences in runs cannot be due to the set of lexicons given

to regular agents, and we can focus our analysis on the lexicons given to IAs. We measured the average distance of IA lexicons from their neighbouring agents' lexicons, and the average specificity of these lexicons, reasoning that these are the two main properties that will affect convention emergence due to lexicon differences. However, we found no statistically significant difference between the runs. Therefore, the only remaining option to explain the differences between the runs is the difference in connecting topology and the placement of IAs on that topology. These results therefore corroborate other results implicating network structure in convention emergence dynamics.

5 Conclusions and future work

Our results show that small groups of unprivileged agents can effectively and significantly influence the emergence of conventions within open MAS. When agents are able to generate high quality conventions to spread to the population, we find that just 5 randomly placed IAs in a population of 1000 can influence the rest of the population to use their conventions 43% of the time. When we place each IA according to highest node degree, we can influence the rest of the population to use the IA convention 100% of the time with only 10 agents in a population of 1000.

Our results with IAs with poor quality lexicons shows that this influence goes both ways, such that convention emergence can be fragmented and almost entirely eliminated with the same small proportion of IAs who use poor quality conventions. We note that these results also imply topological influences have a major effect on the emergence of conventions. Scale-free networks are particularly conducive to high quality convention emergence, but our results suggest that the cost of convention manipulation on small-world networks is higher, and convergence is much slower. It may be possible to place IAs at highly influential locations to increase the probability of convention emergence. In future work, we intend to further investigate the effect of location, as well as exploring the robustness and stability of convention emergence in the presence of agents that do not adopt conventions as described above, or conflict with the rest of the population in some other manner (e.g. by having conflicting goals).

The strategy investigated for IAs in this paper is very simple, namely that of attempting to propagate a single high quality convention. It may be possible to imbue IAs with more complex strategies, including incentives and sanctions for adherence of norms, information propagation, or adaptation to observed conventions in the population. Coordination between IAs, especially in conjunction with topological rewiring, may also significantly increase the efficacy of IAs.

Our results show that small proportions of agents in a population are able to significantly manipulate the conventions that are adopted by relatively large societies, without relying on imposing additional architectural requirements at either agent or society level. However, this manipulation can both impede and aid high-quality convention emergence, and care must be taken with both the strategy that IAs use and their location in the social network. Our results imply that malicious or faulty agents can disrupt convention emergence with ease, and demonstrate the fragility of convention emergence in open MAS. That malicious agents could use dishonest quality valuations or attempt to block the propagation of high quality convention seeds in order to fragment convention emergence conflicts with our desire to minimise intrusive additions or impositions on agent behaviour, especially as such action is typically dealt with either through self-protection (as with Salazar et al. [31]) or the use of social constructs such as sanctions or incentives. We will likely therefore be forced to

compromise on non-intrusiveness to deal with the practicalities of malicious behaviour but we still do not require assumptions about the form this self-protection will take for the above results to hold.

Finally, we note our plans for future work. In this paper, we have assumed a static topology, which is an idealised situation given dynamic real-world application domains such as MANETs or P2P networks. We intend to evaluate the efficacy of our IA mechanism on time-varying topologies, such as the particle collision model for dynamic scale-free networks developed by Gonzalez et al. [12]. We have studied convention emergence in a single abstract model. To confirm the effectiveness of IAs in general will require exploration of the mechanism in other models. However, given the findings in other works discussed in Sect. 2 (e.g. [23]), we anticipate that the IA mechanism will generalise successfully. Finally, although we have shown that the location of IAs is an important factor, developing a full understanding of topological issues remains a task for ongoing investigations. Allowing agents to rewire their connections has been shown to be effective in other domains (e.g. [13]) and given the results regarding targeting of IA location, we expect an IA rewiring strategy would significantly increase their efficacy.

References

1. Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics, 74*, 47–97.
2. Agotnes, T., Vander Hoek, W., & Wooldridge, M. (2009). Robust normative systems and a logic of norm compliance. *Logic Journal of IGPL, 18*(1), 4–30. doi:10.1093/jigpal/jzp070.
3. Arthur, W. (1994). Inductive reasoning and bounded rationality. *The American Economic Review, 84*(2), 406–411.
4. Axelrod, R. (1986). An evolutionary approach to norms. *The American Political Science Review, 80*(4), 1095–1111.
5. Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509.
6. Boman, M. (1999). Norms in artificial decision making. *Artificial Intelligence and Law, 7*(1), 17–35.
7. Chen, W., Wang, Y., & Yang, S. (2009). Efficient influence maximization in social networks. In *Proceedings of the 15th ACM international conference on knowledge discovery and data mining* (pp. 199–208).
8. Dellarocas, C., & Klein, M. (1999). Civil agent societies: Tools for inventing open agent-mediated electronic marketplaces. In *Proceedings of the ACM conference on electronic commerce* (pp. 24–39).
9. Eppstein, D., & Wang, J. (2002). A steady state model for graph power laws. In *2nd international workshop on web dynamics*.
10. Galan, J., & Izquierdo, L. (2005). Appearances can be deceiving: Lessons learned re-implementing Axelrod's evolutionary approach to norms. *Journal of Artificial Societies and Social Simulation, 8*(3), 108–111.
11. Garlick, M., & Chli, M. (2009). The effect of social influence and curfews on civil violence. In *Proceedings of the 8th international conference on autonomous agents and multiagent systems* (pp. 1335–1336).
12. Gonzalez, M., Lind, P., & Hermann, H. (2006). Networks based on collisions between mobile agents. *Physica D: Nonlinear Phenomena, 224*(2–4), 137–148.
13. Griffiths, N., & Luck, M. (2010). Changing neighbours: Improving tag-based cooperation. In *Proceedings of the 9th international conference on autonomous agents and multi-agent systems* (pp. 249–256).
14. Grizard, A., Vercouter, L., Stratulat, T., & Muller, G. (2007). A peer-to-peer normative system to achieve social order. In *Coordination, organizations, institutions, and norms in agent systems II* (Vol. 4386, pp. 274–289). Springer.
15. Huynh, T., Jennings, N., & Shadbolt, N. (2006). An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems, 13*(2), 119–154. doi:10.1007/s10458-005-6825-4.
16. Jennings, N. (1993). Commitments and conventions: The foundation of coordination in multi-agent systems. *The Knowledge Engineering Review, 8*(3), 223–250.

17. Kempe, D., & Kleinberg, J. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM international conference on knowledge discovery and data mining* (pp. 137–146).
18. Kittock, J. (2002). Emergence of social conventions in complex networks. *Artificial Intelligence, 141* (1–2), 171–185. doi:10.1016/S0004-3702(02)00262-X.
19. Kleinberg, J. (2000). Navigation in a small world. *Nature, 406*(3), 845.
20. Mahmoud, S., Griffiths, N., Keppens, J., & Luck, M. (2010). The role of mutation in norm emergence. In *Proceedings of the 5th international workshop on normative multi-agent systems* (pp. 23–28).
21. Modgil, S., Faci, N., Meneguzzi, F., Oren, N., Miles, S., & Luck, M. (2009). A framework for monitoring agent-based normative systems. In *Proceedings of the 8th international conference on autonomous agents and multiagent systems* (pp. 153–160).
22. Morales, J., López-Sánchez, M., & Esteva, M. (2011). Using experience to generate new regulations. In *Proceedings of the 22nd international joint conference on artificial intelligence* (pp. 307–312).
23. Mukherjee, P., Sen, S., & Airiau, S. (2008). Norm emergence under constrained interactions in diverse societies. In *Proceedings of the 7th international joint conference on autonomous agents and multi-agent systems* (pp. 779–786).
24. Oh, J., & Smith, S. (2008). A few good agents: multi-agent social learning. In *Proceedings of the 7th international joint conference on autonomous agents and multiagent systems* (pp. 339–346).
25. Oliver, P. (1980). Rewards and punishments as selective incentives for collective action: Theoretical investigations. *American Journal of Sociology, 85*(6), 1356–1375.
26. Page, S. (1997). On incentives and updating in agent based models. *Computational Economics, 10*(1), 67–87.
27. Perreau de Pinninck Bas, A., Sierra, C., & Schorlemmer, M. (2009). A multiagent network for peer norm enforcement. *Autonomous Agents and Multi-Agent Systems, 21*(3), 397–424. doi:10.1007/s10458-009-9107-8.
28. Pirzada, A. A., & Mcdonald, C. (2006). Trust establishment in pure ad-hoc networks. *Wireless Personal Communications, 37*(1–2), 139–168. doi:10.1007/s11277-006-1574-5.
29. Sabater-Mir, J., Paolucci, M., & Conte, R. (2006). Repage: Reputation and image among limited autonomous partners. *Journal of Artificial Societies and Social Simulation, 9*, 2.
30. Salazar, N., Rodriguez-Aguilar, J. A., & Arcos, J. (2010). Robust coordination in large convention spaces. *AI Communications, 23*(4), 357–372.
31. Salazar, N., Rodriguez-Aguilar, J. A., & Arcos, J. L. (2010). Convention emergence through spreading mechanisms. In *Proceedings of the 9th international conference on autonomous agents and multiagent systems* (Vol. 1, pp. 1431–1432).
32. Salazar, N., Rodriguez-aguilar, J. A., & Arcos, J. L. (2010). Robust coordination through spreading mechanisms. Technical Report IIIA-TR-2010-10, Artificial Intelligence Research Institute, Spanish National Research Council.
33. Salazar, N., Rodriguez-aguilar, J. A., & Arcos, J. L. (2008). Infection-based self-configuration in agent societies. In *Proceedings of the 2008 conference companion on genetic and evolutionary computation* (pp. 1945–1951).
34. Savarimuthu, B., Cranefield, S., & Purvis, M. (2007). Norm emergence in agent societies formed by dynamically changing networks. In *Proceedings of the 2007 IEEE/WIC/ACM international conference on intelligent agent technology* (pp. 464–470). doi:10.1109/IAT.2007.76.
35. Sen, S., & Airiau, S. (2007). Emergence of norms through social learning. In *Proceedings of the twentieth international joint conference on artificial intelligence* (pp. 1507–1512).
36. Sethi, R., & Somanathan, E. (2005). Norm compliance and strong reciprocity. In *Moral sentiments and material interests: The foundations of cooperation in economic life* (pp. 229–250). Cambridge: MIT Press.
37. Shoham, Y., & Tennenholtz, M. (1997). On the emergence of social conventions: Modeling, analysis, and simulations. *Artificial Intelligence, 94*(1–2), 139–166. doi:10.1016/S0004-3702(97)00028-3.
38. Singh, M. (2000). A social semantics for agent communication languages. In *Issues in agent communication* (pp. 31–45). Berlin: Springer.
39. Sommerfeld, R. D., Krambeck, H. J., Semmann, D., & Milinski, M. (2007). Gossip as an alternative for direct observation in games of indirect reciprocity. *Proceedings of the National Academy of Sciences of the United States of America, 104*(44), 17435–17440. doi:10.1073/pnas.0704598104.
40. Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life, 2*(3), 319–392.
41. Szabó, G., & Fath, G. (2007). Evolutionary games on graphs. *Physics Reports, 446*(4–6), 97–216.
42. Vogiatzis, G., MacGillivray, I., & Chli, M. (2010). A probabilistic model for trust and reputation. In *Proceedings of the 9th international conference on autonomous agents and multiagent systems* (pp. 225–232).

43. Walker, A., & Wooldridge, M. (1995). Understanding the emergence of conventions in multi-agent systems. In *Proceedings of the 1st international conference on multi-agent systems* (pp. 384–389).
44. Yu, C., Werfel, J., & Nagpal, R. (2010). Collective decision-making in multi-agent systems by implicit leadership. In *Proceedings of the 9th international conference on autonomous agents and multiagent systems*, Richland, SC (pp 1189–1196).