Overcoming the Tragedy of the Commune in the Hawk-Dove game through Conventional Coding

Joachim De Beule JOACHIM@AI.VUB.AC.BE
Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

Keywords: mechanisms of evolution, language, learning, biosemiotics, multi-agent modeling

Abstract

In the Hawk-Dove game, where two individuals compete over a resource, fully cooperating or "dove"-like behavior is vulnerable to invasion by defecting or "hawk"-like behavior, a fact also known as the "tragedy of the commune". This tragedy can be overcome by so-called "bourgeois" or "anarchist" players which conventionally base their behavior on an external sign. However, it is not a-priori clear how such behavior could evolve through natural selection alone. In this paper it is shown, through simulations, that it can be the result of a more general strategy by which adaptive agents learn and establish a globally shared conventional code.

1. Introduction

The Hawk-Dove game (Maynard Smith & Price, 1973), also known as the snowdrift game or the chicken game, is used to study a variety of topics, from the evolution of cooperation (Doebeli et al., 2004) to nuclear brinkmanship (Russell, 1959). In the game, two players compete over a resource. They can choose between two actions named 'hawk' and 'dove'. If both players play dove then they share the resource. If both play hawk then they share the resource minus a fighting cost. Hawks receive the complete resource when playing against doves. Doves can only thrive if the fighting cost exceeds the reward. In this case they have the advantage over hawks that they can share resources without risking a fight. But hawks can also take advantage of them. In sum, naively cooperative doves are destined to be exploited by aggressive hawks. This 'paradox of cooperation' was termed the 'tragedy of the commune' by Doebeli.

The game can be extended with an uncorrelated asymmetry by informing players about whether they are the first or the second player. If players have equal chances of being first, then they may decide to play hawk when first and dove when second (or vice versa). This strategy was called the 'bourgeois' (or 'anarchist') strategy by Maynard Smith, and is an evolutionary stable strategy (it outperforms hawks and doves). This shows that full cooperation (in the sense of the Hawk-Dove game) can evolve if players can rely on a fair (but otherwise arbitrary) external sign.

It is another question how the bourgeois strategy could evolve. This crucially depends on what it means to be the first or the second player. An obvious choice is arriving first at a resource or territory. For example, territorial disputes between male speckled wood butterflies (*Pararge aegeria*) in England are resolved according to the bourgeois strategy (Davies, 1978). In this case, it is conceivable that the behavior is encoded in the hereditary material and the result of a mutation.[1] This is much less so for other arbitrary signs. Consider for instance a label displaying the letters 'private property'. The meaning of such a sign cannot be genetically encoded, it is cultural. This means that it needs to be learned by new (offspring) players from other (parent) players. Since learning can be costly, it is not a priori clear if this is possible in an evolutionary context. Furthermore, learning presupposes that a stable cultural convention is established and available for learning, which requires further explanation.

In this paper it is shown that 'bourgeois-like' behavior can evolve as a result of a more general 'coding' strategy which allows agents to couple meanings (actions) to signals through learning. This strategy is

Appearing in *Proceedings of BeneLearn and PMLS 2012*.

[1] Although the same butterflies were reported to behave differently at other locations.

more general in the sense that it also supports 'anarchistic' behavior, or any of the other (pure) strategies. It is also more general because it works for any signal, not only for those designating ownership or being first. Therefore it does not require the signals to be encoded in the hereditary material, but only the capacity to couple arbitrary signals to meanings, together with an appropriate learning strategy. As is discussed in this paper, under certain conditions such a capacity allows coding players to beat doves and hawks by establishing a cultural convention by which they can coordinate their actions as a species.

2. The Hawk-Dove game

The payoff matrix for the Hawk-Dove game as considered in this paper is given in table 1. The symbols r and c stand for the resource value and fighting cost respectively, with $r \leq c$. These are the same in every game. Each row (column) in the table corresponds to a possible action of the first (second) player. For example, the first entry in the first row and column indicates that if both players play 'hawk', then each of them receives a payoff of $b + (r - c)/2$. If however the second player plays 'dove', then the first player receives the payoff $b + r$ (first row, second column) whereas the second player only receives a payoff b (second row, first column) etc. In the remainder of the paper, the offset value b is set to $-(r - c)/2$ so that all payoffs are non-negative.

	hawk	dove
hawk	$b + (r - c)/2$	$b + r$
dove	b	$b + r/2$

Table 1. Payoff matrix for the Hawk-Dove game. The possible actions of the first player are listed in the first column of the table. This player will receive a payoff as specified in the table for each possible action of the second player.

In this section, three types of players (strategies) are considered: hawks (always playing the action hawk), doves (always playing the action dove) and bourgeois (play hawk when first, dove when second). The expected payoff for a player of a particular type depends on the strategies used by his opponents in games. Let p_h, p_d and p_b be the fraction of hawks, doves and bourgeois in the population respectively.[2] It is assumed that these fractions correspond to the probabilities to meet an opponent of each type. If all players have equal chances of being first then the expected payoffs per game for each strategy are:

$$\begin{aligned}
\langle \mu_d \rangle &= b + (1 - p_h + p_d)r/4 \\
\langle \mu_h \rangle &= b + (3 - p_h + p_d)r/4 - (1 + p_h - p_d)c/4 \\
\langle \mu_b \rangle &= b + (2 - p_h + p_d)r/4 - p_h c/2
\end{aligned}$$

By setting $p_d = 1 - p_h$ in these expressions, one obtains the standard result that neither the hawk nor the dove strategy are evolutionary stable strategies (ESS's) since both of them can be invaded by the other until the equilibrium ratio $p_h = r/c$ is reached. Both strategies can also be invaded by the bourgeois strategy, which is an ESS with respect to the others. This can also be seen from the phase plot of the replicator system that follows from the expected payoffs as shown in Figure 1 for $r = 2$, $c = 3$ (the offset parameter b only affects the time scale of the dynamics and hence does not influence the plot). The state in which there are only bourgeois type players is marked as '$P_b = 1$'. It is the only asymptotically stable state.

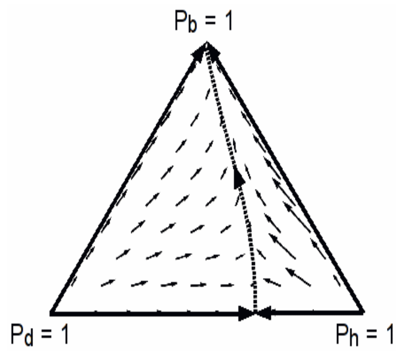

Figure 1. Phase plot of the Hawk-Dove-Bourgeois game replicator system for $r = 2$ and $c = 3$. Each point in the triangle corresponds to specific values for p_h, p_d and p_b. These values are proportional to the distance of the point to the left, right and bottom sides of the triangle respectively. The arrows indicate the direction of the dynamics of the replicator system in phase space.

3. The Coding Strategy and The Effect of Population Turnover

From the previous it follows that if natural selection acts upon a mixture of hawk, dove and bourgeois players then the bourgeois strategy prevails. This result, however, is based on the assumption that offspring of bourgeois players automatically use the same convention, that is, play hawk when first and dove when second. In other words, it assumes that the bourgeois strategy is encoded in the hereditary material of players. It is not obvious how this could be established for

[2]Since these are the only strategies considered it holds that $p_h + p_d + p_b = 1$.

arbitrary and external signals. Investigating the consequences of relaxing this assumption also is interesting in its own right. This is the topic of the remainder of the paper.

If the convention is not passed on to offspring through the hereditary material, then it must be learned. We therefore introduce a new strategy, the 'coding strategy'. Players endowed with this strategy have the *capacity to code* (Barbieri, 2008a), particularly to connect arbitrary signals (percepts) to meanings (responses) through learning. We consider *Roth-Erev* learning with discounting (Roth & Erev, 1995) because it is simple and intuitive, and because it was previously shown to perform well in similar tasks (Catteeuw et al., 2011).

Concretely, the set of possible signals is {F,S} (for First and Second) and the available responses or actions are {H, D} (for Hawk and Dove). For each possible combination $\langle s, a\rangle$ of a signal s and an action a, coding agents keep a score $\phi(s, a)$, initially set to a fixed (genetically encoded) value ϕ_i. When an agent receives signal s^* in a game, it chooses action a^* from all available actions with probability proportional to $\phi(s^*, a^*)$. If after the game the agent receives payoff μ, scores are updated as follows:

$$\begin{aligned} \phi(s^*, a^*) &\leftarrow \lambda\phi(s^*, a^*) + \mu, \\ \phi(s^*, a \neq a^*) &\leftarrow \lambda\phi(s^*, a), \\ \phi(s \neq s^*, a) &\leftarrow \phi(s, a), \end{aligned}$$

where the parameter $\lambda > 0$ is a discounting factor which is also encoded in the hereditary material of coding agents.

Initially, coding agents explore all different actions with equal probability. As they play more games and apply the above update rules, the score of one action per signal will approach the real expected payoff for that action, while all other scores approach zero. Note that the action upon which the agent converges is not necessarily the one yielding the highest expected payoff, although the probability that it is approaches one when the initial score ϕ_i approaches infinity and/or the discounting factor λ approaches zero. However, the time to converge upon a deterministic behavior then approaches infinity as well, so that there is a trade-off between 'exploration' and 'exploitation' (see also Figure 2). This trade-off puts a limit on the capacity of individual agents to learn the optimal behavior. In consequence, just by playing games, a population consisting solely of coding agents will not necessarily be able to establish a fully shared convention such as all agents deterministically playing the bourgeois strategy. This is because, as agents cool down or "grow old", there is always a probability that they settle on different strategies.

It is known that children play a crucial role in the establishment of new natural languages (see e.g. (Senghas & Coppola, 2001; Verhoef & de Boer, 2011)). If a population turnover is added to the model, that is if "old" coding agents are replaced by new ones, exploration continues. Furthermore, new agents will tend to pick up and hence reinforce emerging conventions. This might allow a population of coding agents to establish a shared convention after all. The effectiveness of this mechanism was confirmed through simulations of which the results are shown in Figures 3 and 4. Note that these results depend on the rate at which agents are replaced in the population. This is investigated in more detail in the following.

4. Selection

Selection is brought into the model by replacing agents based on their success in games instead of randomly. We consider a finite population of $N = 100$ agents undergoing natural selection according to a Moran process (Moran, 1962; Nowak, 2006). Each simulation run, the population is initialized to contain a variable fraction $R \in \{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ of coding agents. The rest of the population consists of $100(1 - R)r/c$ hawks and $100(1 - R)(1 - r/c)$ doves, with $r = 2$ and $c = 3$. Between each replacement step, agents are randomly selected to play a total of τ games, with $\tau \in \{500, 1000, 2000, 4000\}$ (that is, on average, each agent plays $2\tau/N$ games before a replacement occurs).

Simulations were run for $\lambda \in \{0.99, 0.95, 0.9, 0.8\}$ and $\phi_i \in \{1, 10, 100\}$. For each combination of parameter values, 100 simulation runs were performed, and the fraction of runs that leads to a population consisting exclusively of coders after a maximum of $2\tau 10^4$ games was recorded. Results are shown in Figure 5. From the figure, it can be seen that coding agents approach the optimal behavior as defined by the behavior exhibited by bourgeois (or anarchist) agents under a wide range of conditions, provided that agents have enough time to learn between successive replacement or selection steps. Only when learning is too greedy (corresponding to low values of λ and ϕ_i –towards the top right of the figure) coders loose their selective advantage over hawks and doves. Still coders experience no *dis*advantage either, as in these cases the probability that they take over the population simply approaches the initial relative abundance R.

5. Discussion and Conclusion

Although it has long been known that the tragedy of the commune can be overcome by players adopting a conventional strategy such as 'bourgeois' and 'anarchist' players, to our knowledge it was not investigated before how such strategies could evolve. The standard explanation is differential reproduction (mutation and selection). This implies that the convention is encoded in the hereditary material. However, because a convention, like language and other cultural phenomena, is *arbitrary*, it is not a-priori clear how this is possible. It was shown that the tragedy of the commune can also be overcome by "coding" players. Coding players learn association strengths between signals and actions that reflect expected payoffs and act accordingly. Interaction between such players induces a positive feedback loop between their preferences. This results in the amplification of –otherwise arbitrary– preferences at the population level, until a globally shared convention emerges with which coding players can coordinate their actions in a nearly-optimal way. In line with what is proposed in (Barbieri, 2008b), I propose to refer to this mechanism of evolution as *evolution by natural conventionalization.*

Crucially, it is the interplay between evolution by natural selection and evolution by natural conventionalization that determines the outcome of evolution in total. Without differential reproduction, that is even without population turnover, evolution stops and no conventionalization takes place. On the other hand, without conventionalization, the Major Transitions in macroevolution might never have occurred. These transitions are characterized by increased degrees of coordination, for instance between cellular agents that became organized into multi-celled organisms (Maynard-Smith & Szathmáry, 1995). As in the Hawk-Dove game, coordination requires that a conventional code is available, and hence that conventionalization mechanisms are at work. It should therefore come as no surprise that the Major Transitions are all accompanied by the appearance of new, arbitrary codes, an example of which is the genetic code (Maynard-Smith & Szathmáry, 1995; Barbieri, 1998).

Like Barbieri, I conclude that natural selection and natural conventionalization are complementary mechanisms of evolution, the first accounting for the gradual transformation of existing species through differential reproduction and the second for the origin and fixation of absolute novelties at higher levels of organization. These mechanisms are not independent however, and their interplay must be taken into account in order to obtain a full understanding of evolution.

References

Barbieri, M. (1998). The organic codes. the basic mechanism of macroevolution. *Rivista di Biologia-Biology Forum*, *91*, 481514.

Barbieri, M. (2008a). Biosemiotics: a new understanding of life. *Naturwissenschaften 95(7)*, 577–599.

Barbieri, M. (2008b). The mechanisms of evolution. natural selection and natural conventions. In M. Barbieri (Ed.), *The codes of life*, chapter Dordrecht, 15–35. Springer.

Catteeuw, D., De Beule, J., & Manderick, B. (2011). Roth-erev learning in signaling and language games. *Proceedings of the 23rd Benelux Conference on Artificial Intelligence* (pp. 65–74).

Davies, N. (1978). Territorial defence in the speckled wood butterfly (pararge aegeria): the resident always wins. *Anim. Behav.*, *26*, 138–147.

Doebeli, M., Hauert, C., & Killingback, T. (2004). The evolutionary origin of cooperators and defectors. *Science*, *306*, 859862. doi:10.1126/science.1101456.

Maynard Smith, J., & Price, G. (1973). The logic of animal conflict. *Nature*, *246*, 1518.

Maynard-Smith, J., & Szathmáry, E. (1995). *The major transitions in evolution.* Oxford: Oxford University Press.

Moran, P. A. P. (1962). *The statistical processes of evolutionary theory.* Oxford: Clarendon Press.

Nowak, M. A. (2006). *Evolutionary dynamics: Exploring the equations of life.* Harvard University Press.

Roth, A. E., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, *8*, 154–212.

Russell, B. (1959). *Common sense and nuclear warfare.* London: George Allen and Unwin. ISBN 0041720032.

Senghas, A., & Coppola, M. (2001). Children creating language: Jow nicaraguan sign language acquired a spatial grammar. *Psychological Science*, *12*, 537–558.

Verhoef, T., & de Boer, B. (2011). Language acquisition age effects and their role in the preservation and change of communication systems. *Linguistics in Amsterdam*, *4*.

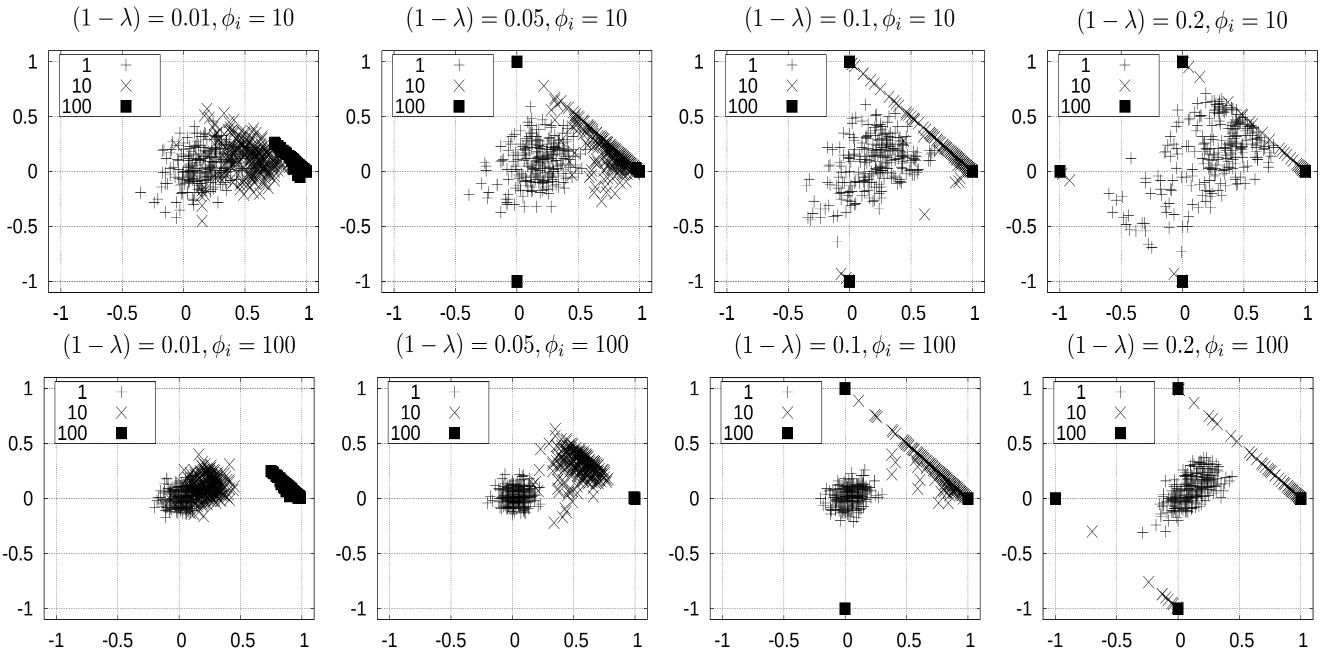

Figure 2. Sampled behavior of coding agents learning by playing games with dove agents. The point at $(x = 1, y = 0)$ corresponds to pure hawk-like behavior. The opposite point at $(x = -1, y = 0)$ to pure dove-like behavior. The points at $(x = 0, y = 1)$ and $(x = 0, y = -1)$ correspond to pure bourgeois- and anarchist-like behavior respectively. Mixed behaviors are also possible since coder agents are not necessarily deterministic. The different point types correspond to the amount of games played. The more games played, and depending on the learning parameters λ and ϕ_i, the more deterministic the behavior becomes. The optimal behavior in this case is hawk-like, which is eventually reached when learning parameters are "favorable", that is towards the left and bottom of the Figure.

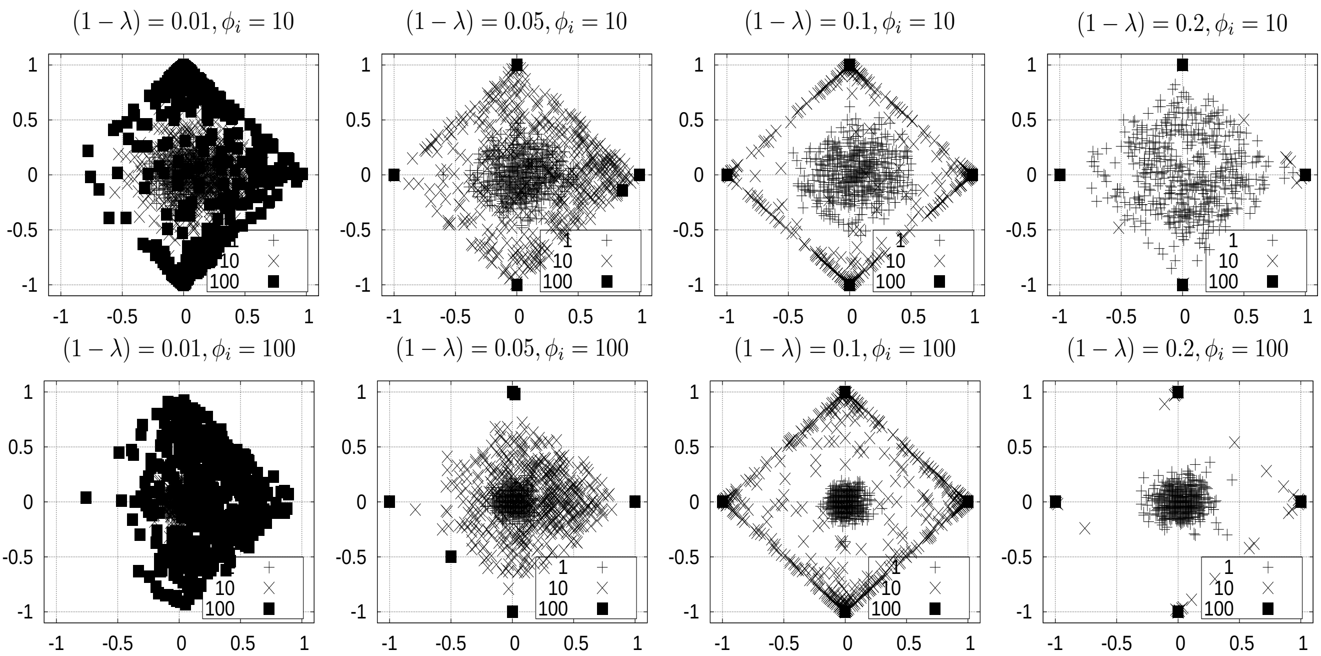

Figure 3. Same as in Figure 2 but instead of playing against doves, the coding agents now learn by playing against other (learning) coding agents. Due to the exploration/exploitation trade-off, a convention is not always established, that is the agents do not always adopt the bourgeois or anarchist strategies, not even when learning parameters are "favorable" (see fig. 2)

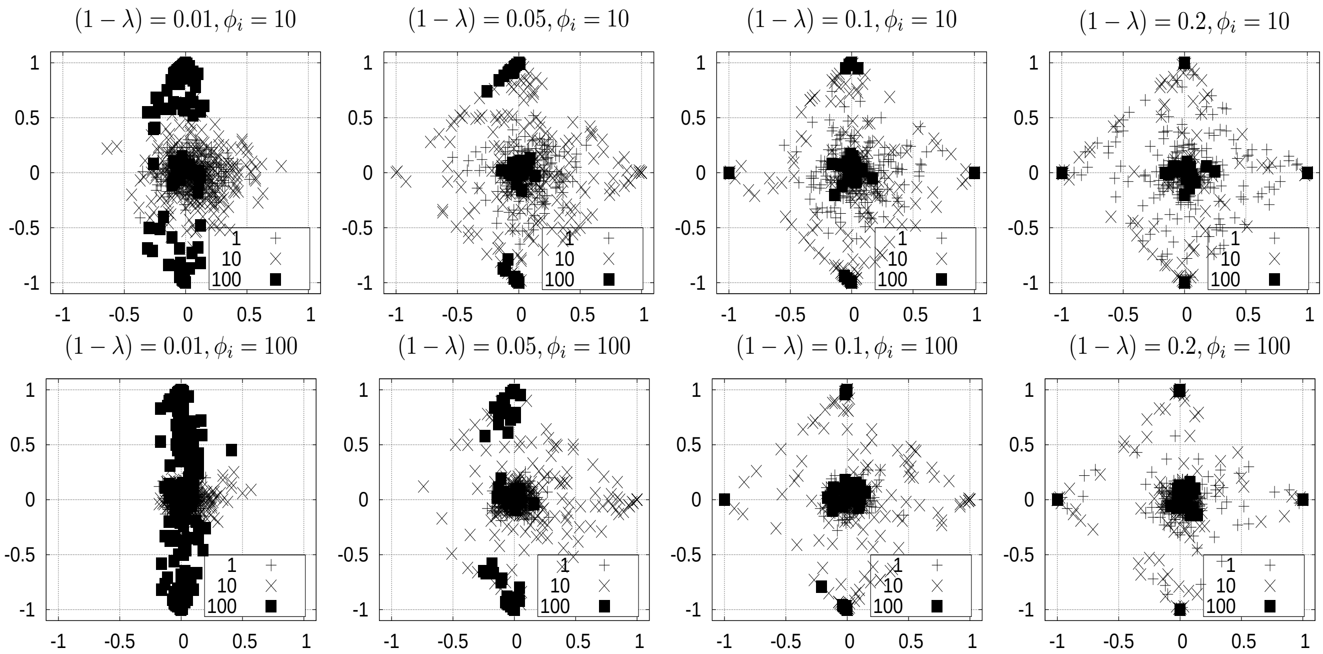

Figure 4. Same as in Figure 3, but now agents are randomly replaced in the population at some fixed rate. If learning parameters are favorable, a convention emerges and agents eventually adopt the bourgeois or anarchist strategy. The reason why convergence never appears to be complete is because new agents are constantly entering in the population and because these still need to adopt the convention.

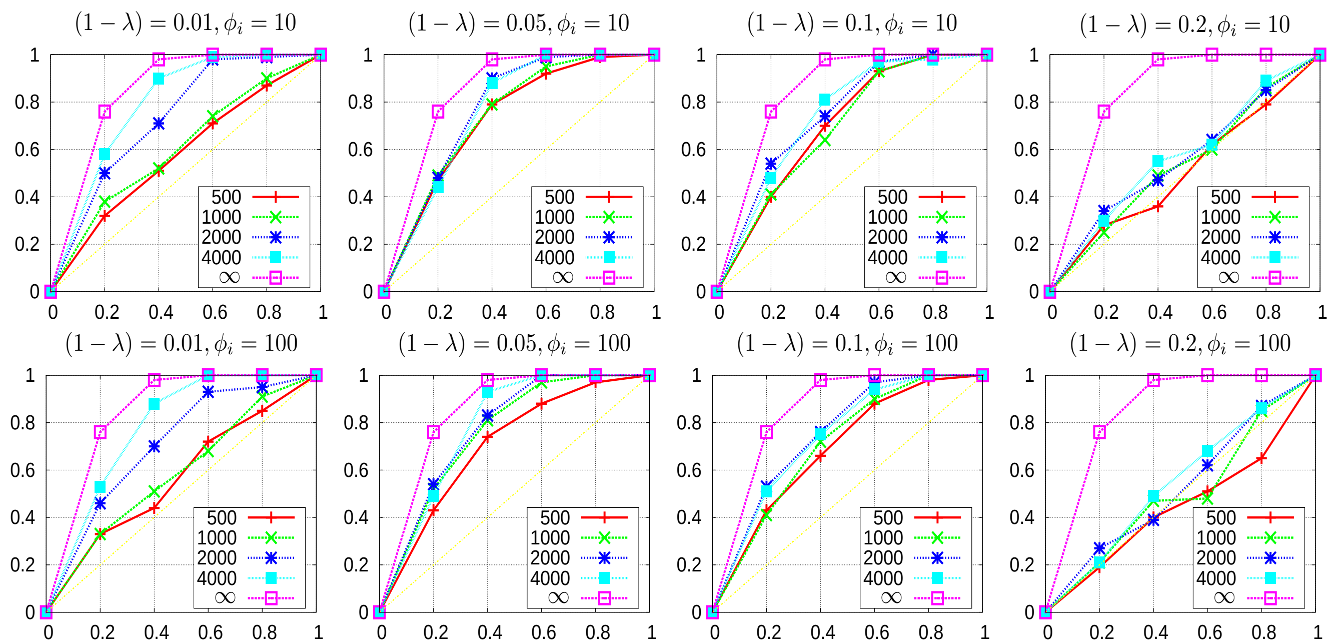

Figure 5. Fraction of simulation runs that lead to an end state of only coders for different values of the learning parameters λ and ϕ_i. The X-axis indicates the initial fraction of coding agents R. Values that are above the bisecting line 'y=x' indicate a selective advantage for coding agents over hawks and doves. Each curve is labeled with the number of games τ in between two successive (Moran) replacement steps. The curves labeled '∞' were obtained with bourgeois agents instead of coders, and indicate the best attainable behavior. Coding agents approach this optimal behavior if learning parameters are favorable *and* if there is enough time to learn between successive replacement events (τ large). The learning parameters determine the speed of learning which in turn interferes with how much time is available for learning, which explains why the optimal set of parameters now is no longer found simply towards the bottom left of the Figure.