Investigating Convention Shifts and Team Reasoning in Multi-Agent Simulations

Anna Coenen (anna.coenen.09@ucl.ac.uk)
Ramsey Raafat (r.raafat@ucl.ac.uk)
University College London, Department of Psychology
26 Bedford Way, WC1H 3AE London, UK

Nick Chater (Nick.Chater@wbs.ac.uk)
Warwick Business School
CV4 7AL Coventry, UK

Abstract

Classical game theory struggles to explain how rational players should decide between a number of social conventions, even if some yield higher individual payoffs than others. Thus, on a population level a group or society may be stuck in using one convention when there exist alternative and potentially more beneficial ones. Using an agent-based model the current study examines how convention shifts from less to more beneficial conventions can come about. To investigate this, we use the concept of team reasoning, a mode of reasoning in which actors maximise the utility of a group rather than their own. Unlike other social decision-making mechanisms, such as forms of imitation, team reasoning enables the spread of a more profitable convention through a population even if no global knowledge about the population is available to agents.

Keywords: Conventions; Team Reasoning; Agent-Based Simulation; Equilibrium Selection; Social Learning.

Introduction

Social conventions enable people to act in situations that require coordination with others. That is, situations in which all participants can profit if they follow the same course of behaviour or the same set of rules. Examples of social conventions that regulate our daily interactions are traffic rules, language, currencies, or property. More recently, the internet has created new kinds of interaction which crucially rely on social conventions: Successful use of social networks, for instance, primarily relies on other people using the *same* service, and only secondarily on the potential advantages one service may have over others.

A central property of social conventions is that they are *self stabilising* and *self perpetuating* (Lewis, 1969). Once a convention has been established and it yields a minimal benefit for everyone, it will be difficult to abolish or alter. This article will examine a special case of this problem: Given that a group of people, or a society has converged on using one particular convention, how can it change this convention or start adopting an overall more beneficial one?

To investigate this question we will study conventions as outcomes of *coordination games*, which will be described in the next section.

Conventions in Coordination Games

In classical game theory, conventions can be expressed as equilibria in coordination games, such as the Stag Hunt game or the Hi-Lo game (see table 1). These are games with several strict Nash Equilibria that require players to select corresponding strategies; in most cases this means players have to choose the same action in order to receive the highest payoff from each strategy. If these equilibria or conventions can also be pareto-ranked, coordination games illustrate the dilemma that players can get 'stuck' using one convention even if a more beneficial alternative exists. This is what happens when players converge to a pareto-inferior Nash Equilibrium (such as (C1, C1) table 1) that yields lower payoffs than another equilibrium in the game for at least one of the players and does not improve any other player's outcome. Although it seems intuitive for each player to aim for the

Table 1: A Hi-Lo Coordination Game

P1/P2	C1	C2
C1	2, 2	0, 0
C2	0, 0	5, 5

pareto-dominant equilibrium when choosing their strategies (Harsanyi & Selten, 1988), according to classical game theory players have no rational reason to prefer it over inferior alternatives (Bacharach, 2006). This is because each player only has a reason to play any equilibrium strategy if the other player(s) do so as well, which, in turn, again depends on this player's strategy choice, etc., leading to an infinite regress of the strategy selection process. Coordination games are therefore an instance of the more general *Problem of Equilibrium Selection* within game theory. To investigate how it is possible for a group or society to overcome this problem and to converge to playing payoff-dominant conventions is the aim of this study.

Note, that the type of equilibrium selection problem in coordination games differs fundamentally from that of another family of well-studied problems: games of cooperation. An example of the latter is the famous Prisoner's Dilemma in which players can profit from cooperation, but can also exploit each other's willingness to cooperate. Cooperation games are therefore particularly relevant for the study of behavioural norms and moral rules that may require being enforced by forms of punishment against defectors. Unlike these games, coordination problems do not lead to a conflict between the collective and self-interest of players. Instead, they show how groups of rational players with similar interests can end up in a state that is undesirable for all.

Solving the Problem of Equilibrium Selection

The equilibrium selection problem in coordination games has triggered a range of responses, both from psychologists and game theorists. Schelling (1960) for instance famously showed how people often select equilibria according to how salient their respective labels are. According to this approach successful coordination relies on shared representations between players. Other authors proposed that behaviour in coordination games can be modelled using some form of multiple-level reasoning in which players select their strategies based on their assumptions about the belief-state of the other player (e.g., Crawford, Gneezy, & Rottenstreich, 2008; Camerer, Ho, & Chong, 2004). Such a theory runs into similar problem as classical game theory, however. Again, an agent's assumptions about its co-player's beliefs rely on what the latter beliefs what the former beliefs, etc. Thus, without any strong initial reason to believe that another player would play the payoff-dominant option, this process leads to a similar type of infinite regress that lies at the heart of the equilibrium selection problem.

Social Learning   Whereas such theories approach the problem of equilibrium selection on an individualistic level, it may be more useful to tackle the question of how convention changes come about on a population level directly. Within the literature on social learning, authors have more generally addressed the problem how norms and conventions can change and stay adaptive given that they are transmitted between individuals. It has been suggested that humans can weed out maladaptive traits or behaviours via selective imitation mechanisms that rely on the success of a certain trait or norm (Laland, 2004; Boyd & Richerson, 2005). By imitating successful individuals in particular, a population can pick up new and more beneficial behaviours, rather than keep transmitting out-dated information by blind copying. Selective imitation mechanisms are thus a candidate for the driving force of convention changes.

Team Reasoning   Classical game theory, Schelling's focal point approach, as well as the social learning literature are grounded in the assumption that players choose the strategies that best promote their individual payoff. This individualistic assumption is transcended by the team reasoning approach (e.g., Gold & Sugden, 2007; Sugden, 2000). Team reasoning assumes that, in interactive situations, people often maximise the payoff of a group or team, rather than their own. Once several players have established their membership to the same group, they can play their part of a *joint strategy* which maximises the payoff of all players combined. The theory presupposes that players enter this special mode of reasoning once they have established their membership to a team. This can happen via explicit agreement but also due to shared experiences or mutual observations (Sugden, 2003).

Team reasoning points to some important features of human interactions that can help solve the problem of equilibrium selection in coordination games. Given that they manage to reach mutual confidence in some way, team reasoners can coordinate their actions through conventions that maximise their collective payoff. Similarly, team reasoning should also be able to motivate convention change. If, in a society or group, there exists a payoff-dominant alternative to a convention and a team of individuals are aware of it, they are expected to behave in accordance with this new convention when interacting with each other the next time.

In this study we explore the prerequisites and dynamics of convention change using an agent-based simulation, asking (i) what the general dynamics are that underlie convention shifts in a population, and (ii) how team reasoning and social learning mechanisms compare in their ability to trigger them.

Study I: Dynamics of convention shifts with simple reinforcement learning

The first study aims to demonstrate the dynamics of a simple version of our model in which agents update their strategies using a reinforcement learning rule. In the second study we will then turn to a strategy comparison between social learning and team reasoning.

The Model

A population of 49 agents is placed in a 7 by 7 lattice that is folded from North to South and East to West. This yields a toroidal structure without any edges and guarantees that all agents have the same number of neighbours on all sides. Each agent can then interact with its eight direct neighbours (to the N, NE, E, SE, S, SW, W, and NW). Agents repeatedly interact with one another in pairs, each time having to coordinate their actions. Each agent can choose from two actions, corresponding to two conventions: Convention 1 and Convention 2 (henceforth: *C1* and *C2*). Subsequently they receive their payoff from this interaction according to the payoff matrix shown in table 1. The payoff structure is that of a pure coordination game with two Nash Equilibria:(C1,C1), in which both players receive a payoff of 2, and (C2,C2), where agents both receive a payoff of 5; if coordination fails, both agents get a payoff of 0. The equilibrium (C2,C2) pareto-dominates (C1,C1) thus representing a case in which one convention, if applied successfully, is more beneficial than the other.

The simple learning rule   When agents interact they decide which strategy to play in the coordination game using a *strategy selection rule*. Previous studies on conventions in agent-based simulations have used some type of reinforcement learning to model strategy selection, that takes into account an agent's history of strategies played and payoffs received (for example, Shoham and Tennenholtz (1997) and Delgado (2002) using a highest cumulative reward rule for strategy selection; Barr (2004) using a one-layer neural network). Similarly, in this first study it is assumed that agents rely on a simple learning rule that chooses a strategy based on the payoff it has generated for that agent in its remembered past. During each encounter both agents look back at their last *m* interactions, where *m* corresponds to an agent's
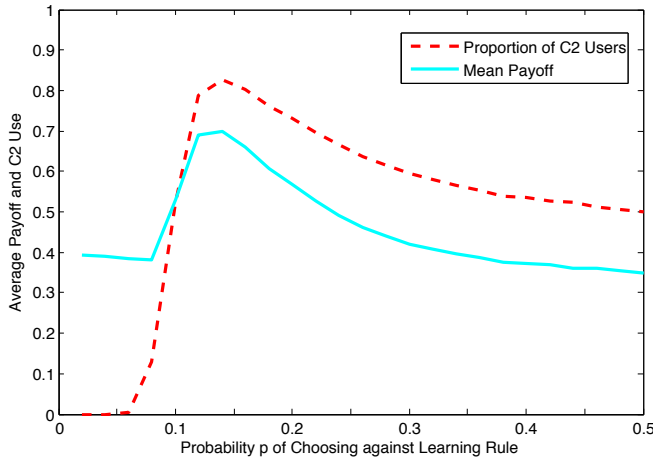
Figure 1: Average payoff and use of C2 for different probabilities p of choosing against the learning rule

Figure 2: Average payoff and use of C2 for different levels of initial C2 users

memory capacity. Then, they choose the strategy that has yielded the highest average payoff over these interactions. If both strategies generated the same average payoff in that time period, agents randomly pick one of them. If one strategy has never been played and therefore not remembered before, its average payoff equals zero. At the beginning of the simulation all agents remember only having played strategy C1, and always having received the associated payoff of the coordination equilibrium (C1, C1).

Given this model, one can now investigate the roles that different key parameters play in shifting the population from using convention C1 to using convention C2. To illustrate, it shall be shown what impact different ways of introducing the better convention have on the dynamics of the model. Unless stated otherwise, all results reported in the next section, were obtained from simulating a population of 20 agents with a memory capacity, $m = 3$ [1] that were randomly matched for 2000 interaction rounds. The outcome measure of each simulation is the proportion of agents that, in their last interaction, adhered to the better convention, C2, and the average payoff all agents received in this interaction. Each simulation was repeated for 1000 runs and the outcome measures reported are averages over these runs.

Results: Basic model with simple learning

One way of introducing the new convention in the population is by letting agents randomly deviate from their strategy choices with a certain probability p. In the initial population this will lead to agents occasionally 'trying out' to act in accordance with the new convention.

Figure 1 shows the average use of strategy C2 in the population for different levels of p. If p is very low (between 0 and 0.06) C2 almost never spreads through the population. Once

---

[1]This memory size could be shown to produce the greatest probability of convention shifts to happen in the current population, all else being equal.
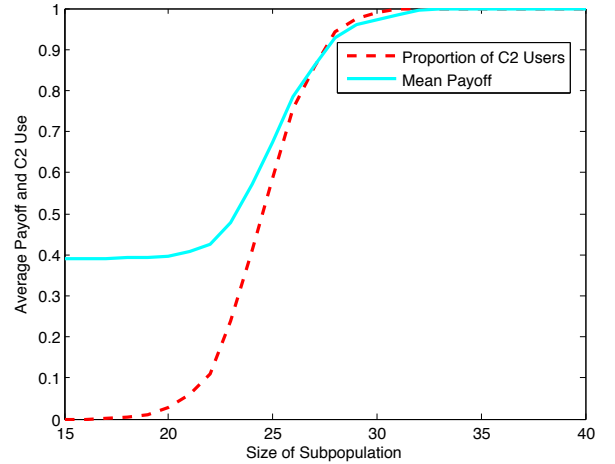
p increases, the amount of times that a population flips from most agents adhering to C1 to most agents adhering to C2 grows rapidly. Note that, because of the random parameter p, no population ever converges to using purely one convention, but for lower levels of p both conventions are basins of attraction. As p approaches 0.5, naturally, the behaviour of the population approaches randomness as agents cannot rely on their learning rule any more. Given the current parameters, this shows that a certain degree of 'adventurousness' can trigger population to shift from one convention to another, payoff dominant one, without communication between the agents.

A second way of introducing a convention in a population, and possibly a more realistic one regarding how new conventions get introduced in real societies, is by having a group of agents enter this population that already adhere to the new convention. Using the same parameter settings as before, but with no random perturbation of strategy use, it is possible to introduce a subpopulation of C2 users in the population. In the beginning of each run, agents in this subpopulation remember only having played strategy C2 and receiving the associated higher payoff from the more beneficial convention. Figure 2 shows the behaviour of the population dependent on the size of this subpopulation.

As expected, the proportion of times a population shifts strictly increases with the size of the subpopulation. Up to a size of 22 initial C2 users, the new convention practically never catches on in the population. Since agents from the new group initially have negative experiences with C2 when interacting with C1-playing agents from the population, they often adapt rather quickly to playing C1 as well. It takes almost the same amount of C2 users compared to C1 user to turn the behaviour of the whole population more than half of the time. This shows just how difficult it is to break with an established convention, even given a significant amount of people that have successfully used a new and better convention already.

Study II: Imitation and team reasoning

Using a basic reinforcement learning rule the model does not seem to explain how it is possible for a new convention to catch on in a group or society given less beneficial conditions, for instance, when a large majority is not already adhering to the new convention. However, in interactive situations people usually do not exclusively use information about their own experience to make a decision, but also consider cues in their social environments, such as the actions and experiences of others. As mentioned above, one important driving-force of cultural evolution are social learning strategies that rely on selective imitation or copying mechanisms (Boyd & Richerson, 2005; Laland, 2004). Moreover, in interactive situations people may be using different modes of reasoning altogether. Thus they might aim at maximising the payoff of a group or team of people, rather than just their own utility.

This second study therefore aims at exploring convention change under more realistic assumptions about the strategy selection process. Three questions shall be investigated: (i) how two well-studied social learning rules and (ii) a version of a *team reasoning rule* manage to explain convention shifts and (iii) how the performance of these different strategy selection rules compares in achieving this. It is of particular interest whether team reasoning can explain convention shifts more successfully than the more widely studied social learning rules. Since the concept sets out to explain how humans derive solutions of coordination problems, it should give some additional insight into the dynamics of convention change.

In the next section, the three new strategy selection rules will be explained. Subsequently, their performance in triggering convention shifts will be investigated in another round of multi-agent simulations.

Strategy Selection Rules

We tested three strategy selection rules:

Imitate-the-best-neighbour (Imitation 1) First, we draw on a well-studied class of mechanisms of social learning are those involving selective copying or imitation of the behaviour of others. One common variant of such imitation learning is copying the behaviour of the most successful member(s) of a group (e.g., Henrich & Gil-White, 2001; Gigerenzer, 2010). In this study we therefore specified an imitation mechanism as follows: When choosing which strategy to play (C1 or C2), an agent determines which of her neighbours received the highest payoff in their last interaction and imitates that strategy. If all neighbours have been equally successful, or have been using the same strategy, the agent uses the simple learning rule specified above.

Imitate-the-best-strategy (Imitation 2) A second imitation rule is to copy not the most successful neighbour, but the strategy that has yielded the highest average payoff in one's neighbourhood (see, e.g., Alexander & Skyrms, 1999). It could be said that this is a more careful version of imitate-the-successful, as it does not rely only on the success of one single neighbour, but on the overall performance of the strategies. As with the first imitation rule, if there is a tie between the strategies' average payoffs or if only one of the strategies has been played in a neighbourhood, agents use their simple learning rule to pick a strategy.

Team Reasoning The concept of team reasoning supersedes the assumption underlying the other two strategy selection rules discussed so far. These assumed that agents rely on some sort of learning to help maximise their individual payoff. Team reasoning, on the other hand, assumes that sometimes what people maximise is the payoff a group or team of people. In order to count towards such a group or team, there has to exist a mutual confidence between members that establishes common knowledge of group membership. Such confidence can be installed for instance by explicit agreement, but also shared experience (Sugden, 2000, 2003). Since the current model assumes no direct communication between agents, group membership is established by shared experience: Given two agents who can observe a successful application of the better convention (i.e. observe two agents using C2), and these two agents can also observe each other, they establish their membership to a team. The next time they interact with each other, they will play their parts in maximising this team's payoff and thus adhere to C2. If there is no common knowledge between them, they resort to the simple learning rule.

Results: Comparing the strategy selection rules

The three strategy selection rules as well as simple learning were pitted against each other using the same model as in the previous study. For the two imitation rules and team reasoning it was assumed that agents could observe the actions and payoffs of their eight neighbouring agents and use this information to update their strategies.

Figure 3 shows the average use of the new convention in a population for different sizes of initial C2 users, separately for each strategy selection rule. Both imitation and team reasoning lead to a higher rate of convention shifts than the simple reinforcement learning rule for most levels of subpopulation size.

For the two imitation rules, this result confirms previous research on social learning, showing that selective imitation can increase the adaptiveness of behaviour. Copying successful strategies thus facilitates the spread of group beneficial conventions, even given that another inferior convention initially prevails in a population. In this study the *imitate-the-best-neighbour* (Imitation 1) rule outperformed the *imitate-the-best-strategy* (Imitation 2) rule slightly for all levels of subpopulation size. This is not surprising given that the former always leads to strategy switches when the latter does, but not vice versa. This is because if one neighbour has applied C2 successfully in their last interaction, that neighbour will be imitated for certain using Imitation 1. With Imitation 2, this one positive interaction might be cancelled out in the
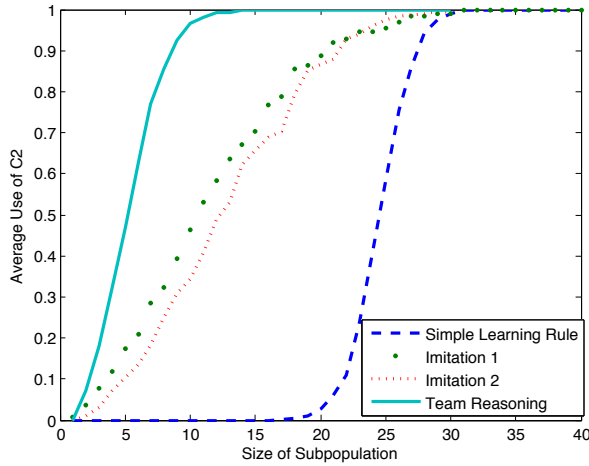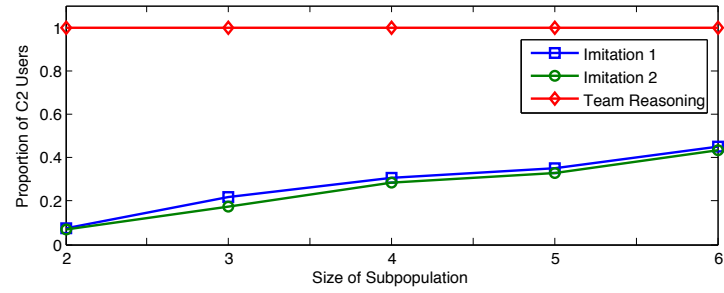
Figure 3: Average use of C2 for different sizes of a subpopulation adhering to C2, by strategy selection rule

average payoff of C2, if two or more other neighbours have failed when using it in their last interactions.
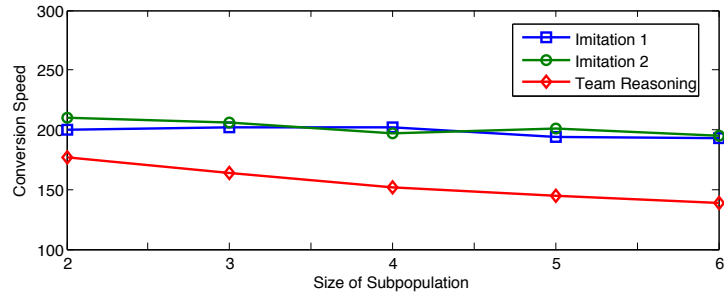
Why is team reasoning so much more successful at triggering convention shifts than the two imitation rules? According to team reasoning, once two agents in a population interact with each other using C2, there will always be surrounding agents observing them and thus establishing mutual confidence. This is why, a single successful interaction can initiate the deterministic spread of the better convention. Whether a population switches conventions, then only depends on the likelihood of two agents of the initial subpopulation meeting each other while they are still adhering to C2.

In contrast, the two imitation rules do not lead to such a deterministic spread of C2. Consider the first imitation rule: If two interacting agents use C2, they will probably be imitated by their surrounding neighbours. However, it is not guaranteed that these neighbours similarly meet agents that imitate the initial successful pair as well, for instance because they are not themselves neighbours of this pair. Hence, applying C2 might fail repeatedly, eventually also causing the initial C2 users to switch to using C1.

To illustrate this difference between the team reasoning rule and the two imitation rules, the model was run again starting with a non-random configuration of agents that belong to the initial subpopulation of C2 users. Two to six subpopulation members were placed in neighbouring cells in the lattice of agents and were the first ones to interact in each run. As a consequence, at the beginning of each simulation there would always be two or more C2 users interacting with each other. Figure 4a shows the proportion of convention shifts resulting from these different starting configurations for the three strategy selection rules. As expected, the team reasoning rule *always* leads to the conversion of the population to the new convention under these parameter settings. This was not the case for the two imitation learning rules, although both performed better than previously when subpopulations of the

(a) Average use of C2

(b) Median conversion speed

Figure 4: Average use of C2 and conversion median speed for clustered subpopulation size, by strategy selection rule

same sizes were not clustered, but randomly distributed in the population.

This difference also manifests itself in the number of interactions that it takes a population to completely converge to the better convention. Figure 4b depicts the median number of interactions it took a population to converge to C2, again given that a subpopulation of C2 users is clustered together in the lattice of agents. The team reasoning rule in general converges faster than the two imitation rules and its conversion speed decreases with the number of agents in the clustered subpopulation. This decrease is not detectable for the two imitation rules. This illustrates that team reasoning leads to the steady spread of a convention through a population, while imitation rules take a more complicated route that involves much 'blind' copying of strategies, leading to repeated failures and slower progress.

Discussion

Using various multi-agent simulations the current study investigated how conventions can change in a population. It was shown that two imitation learning rules and one team reasoning rule for strategy selection could outperform a simple reinforcement learning rule in triggering convention shifts. Moreover, the team reasoning rule proved to be more successful than imitation, providing a possibility of establishing common knowledge and playing joint strategies without central coordination.

The study of conventions is relevant for the question of the adaptiveness of culture. As has been discussed above,

whether culture is adaptive or not crucially depends on the capacity of norms and conventions to transform and adapt to changing environments (Rogers, 1988; Boyd & Richerson, 2001; Enquist, Eriksson, & Ghirlanda, 2007). The current study has confirmed some of the key characteristics authors have attributed to conventions. They are self-stabilising patterns of behaviour that are not easily overturned as long as self-interested agents get some benefit from adhering to them. This is a problem for the claim that culture is adaptive, since it makes it very difficult for a new, more beneficial, convention to replace a current one.

As has been shown here, adaptive filtering mechanisms, such as forms of selective imitation, can aid convention shifts under certain conditions. Such mechanisms have proven to be successful in spreading behaviours whose payoff does not depend on other individuals but rather, for instance, on states in the environment (Enquist & Ghirlanda, 2007; Laland, 2004). However, since conventions rely on the coordination with others, imitation often leads to failures in the early stages when the old convention still prevails. This is the case even if copying is selective, that is, when only successful strategies or individuals are imitated. Thus, our study suggests that imitation may not be among the motors of convention changes, since there is often no immediate advantage of imitating other people's successful behaviour when playing coordination games.

What seems to be necessary for a new convention to replace another one without leading to excessive failures of coordination, is a form of common knowledge between interacting individuals that unites them as a team with common preferences. As proponents of the team reasoning view on strategy selection (Sugden, 2003) have pointed out, people often maximise the utility of a team or group when interacting with others. This is one reason why, for example, they find it easy to choose payoff-dominant equilibria in coordination games. Similarly, the current study has shown that common knowledge helps to promote equilibrium shifts on a population level. A sense of common interest between agents and mutual confidence in group membership, can thus help to change conventions adaptively, and replace current coordinative behaviours with more beneficial alternatives.

References

Alexander, J., & Skyrms, B. (1999). Bargaining with neighbors: is justice contagious? *The Journal of Philosophy*, 588–598.

Bacharach, M. (2006). The hi-lo paradox. In N. Gold & R. Sugden (Eds.), *Beyond individual choice: Teams and frames in game theory.* Princeton University Press.

Barr, D. (2004). Establishing conventional communication systems: Is common knowledge necessary? *Cognitive Science*, *28*(6), 937–962.

Boyd, R., & Richerson, P. (2001). Norms and bounded rationality. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality - the adaptive toolbox* (pp. 281–296). MIT Press.

Boyd, R., & Richerson, P. (2005). *The origin and evolution of cultures*. Oxford University Press, USA.

Camerer, C., Ho, T., & Chong, J. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, *119*(3), 861–898.

Crawford, V., Gneezy, U., & Rottenstreich, Y. (2008). The power of focal points is limited: even minute payoff asymmetry may yield large coordination failures. *The American Economic Review*, *98*(4), 1443–1458.

Delgado, J. (2002). Emergence of social conventions in complex networks. *Artificial intelligence*, *141*(1-2), 171–185.

Enquist, M., Eriksson, K., & Ghirlanda, S. (2007). Critical social learning: A solution to rogers's paradox of nonadaptive culture. *American Anthropologist*, *109*(4), 727–734.

Enquist, M., & Ghirlanda, S. (2007). Evolution of social learning does not explain the origin of human cumulative culture. *Journal of theoretical biology*, *246*(1), 129–135.

Gigerenzer, G. (2010). Moral satisficing: Rethinking moral behavior as bounded rationality. *Topics in Cognitive Science*, *2*(3), 528–554.

Gold, N., & Sugden, R. (2007). Theories of team agency. In F. Peter & H. B. Schmid (Eds.), *Rationality and commitment.* Oxford University Press.

Harsanyi, J., & Selten, R. (1988). A general theory of equilibrium selection in games. *MIT Press Books*, *1*.

Henrich, J., & Gil-White, F. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, *22*(3), 165–196.

Laland, K. (2004). Social learning strategies. *Learning & Behavior*, *32*(1), 4-14.

Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Rogers, A. (1988). Does biology constrain culture? *American Anthropologist*, *90*(4), 819–831.

Schelling, T. (1960). *The strategy of conflict*. Cambridge: Harvard University Press.

Shoham, Y., & Tennenholtz, M. (1997). On the emergence of social conventions: modeling, analysis, and simulations. *Artificial Intelligence*, *94*(1-2), 139–166.

Sugden, R. (2000). Team preferences. *Economics and Philosophy*, *16*(02), 175–204.

Sugden, R. (2003). The logic of team reasoning. *Philosophical Explorations*, *16*(3), 165 - 181.