

Rationality, complexity and self-organisation

Abstract

I discuss the definition of a rational agent in a set of game theoretical scenarios commonly used to study competition and collaboration in social and economic interactions. In particular I analyse the relation between rationality and the ability of a community of agents to self-organise into viable configurations. I suggest that a useful definition of rationality depends on the specific structure of a problem and consequently a common definition which applies to all scenarios is not available. Unless rationality is defined a priori or obtained by induction via an extensive analysis of a given problem, it seems sensible to accept an evolutionary view according to which the concept of rationality is imported into a problem from the experience accumulated in similar settings and modified if evidence requires it.

1 Introduction

One area of interest in complex systems science deals with how agents interact in social and economic problems and how their interaction results in different patterns of organisation. This problem is of current interest because it applies, among other things, to the management of natural resources and thus has the potential to severely affect our environment, the wellbeing of millions of people and the status of several biological species.

The analysis of how people self-organise in different social structures and under different conditions has long drawn the attention of several disciplines from philosophy to history, sociology, economics and evolutionary theory. Much of the views and jargon currently used in complex systems science is obviously influenced by this tradition. Of particular interest to our discussion is the concept of rationality. The word has different meanings and, most important, different connotations, in different fields and has attracted considerable academic discussion. In classic economic theory, an agent acts rationally when it has full knowledge of a problem and employs that knowledge to take a decision which is economically optimal for itself; under certain circumstances, a group of agents all acting rationally will reach a configuration which is also globally optimal, that is they will optimally self-organise.

This view of rationality is questioned mainly for three reasons: first, agents in real world complex problems rarely, if ever, have perfect knowledge of the problem; second, even if they do, evidence shows that they may decide not to follow the strategy suggested by rationality, be this for emotional, ethical or other undetermined reasons; third, there are classes of problems in which following individually rational strategies leads to sub-optimal, at times disastrous, global outcomes; the latter are often referred to as the paradoxes of rationality (Campbell and Sowden, 1985). While the first two points question whether this form of rationality effectively applies to human behaviour, the last point questions the validity of the concept itself.

The methods and approaches of complex systems science are influenced by these concerns; here the term is usually understood as ‘bounded rationality’: it is accepted

that agents never have full knowledge of a problem but, given limited knowledge, still act in order to maximise their individual, and possibly selfish, return. As a result the paradoxes of rationality stand and represent a challenge for our understanding of how economic and social agents self-organise and how they interact with natural resources and biological species.

In this paper I discuss a number of simple game theoretical scenarios which are commonly considered as generalisations of real world problems and employ these scenarios to study the relation between different views of rationality and self-organisation. I suggest that, unless imposed externally by a priori ideology, a unified view of rationality which applies uniformly to all scenarios is not obvious; rather a meaningful definition of rational behaviour depends on the context and structure of the game: when an agent faces a new problem, determining this rational behaviour will inevitably be part of problem solving itself.

Before discussing what we may expect a rational agent to do under different scenarios it may be useful to address the meta-question of what we expect a rational behaviour to involve and why this is important. In this paper I assume an agent to be rational if it has a purpose and takes actions in order to achieve that purpose. In complex system science a definition of rational behaviour is usually employed in order to establish a 'default' action we expect agents to take and then study its possible consequence under different conditions. Since studies of this kind are normally carried out by computer simulation or mathematical analysis, it is inevitable to require that a rational behaviour be described in such terms, i.e., it be algorithmic (it can be reduced to a formula or a set of computer instructions). Since current knowledge and technology prevents us from implementing purely emotional or psychologically motivated decisions in algorithmic form, we are technically prevented from coding them in our models and consequently from considering them as 'rational'; notice however that according to this view random choices can be rational (so far as we accept an algorithmic version of randomness and we disregard deeper mathematical and information theoretical concerns); in other words, agents may decide to let their actions be guided by chance. To summarise, in our discussion a rational behaviour is a set of instructions an agent follows in order to achieve a purpose; the question is what the purpose should be and how the instructions should be given.

2 The basics: the prisoners' dilemma

The prisoners' dilemma (Axelrod, 1984) has been extensively studied in the literature and because of its simplicity it offers a good starting point for our discussion. Among its different possible formulations here I adopt the following. We have two agents; each owns an object that it values at \$1; each values the other agent's object at \$2; they agree to swap the objects; if they do, each will then own an object which it values at \$2 and each will have increased its wealth. However, the transaction is carried out in such a way that each agent may cheat: it may deliver an empty box, not containing the object. If this happens the cheater will obtain the desired object and keep its own for a total value of \$3, while the cheated agent will be left with nothing.

The mainstream view of the problem is that it is rational for each agent to cheat. Suppose you have to decide whether or not to place your object in the box just before the transaction; you may think that the other agent is *independently* facing the same decision; if it decides not to include its object, thereby trying to cheat you, you are better off keeping your own object to avoid being cheated; if it decides to include its object, you are still better off keeping your own, thereby cheating, and ending up with both yours and its object. In both cases, independently of what the other agent does, you appear to be better off by not fulfilling your part of the agreement. However, if both agents follow this rational choice, they both fail to carry out the swap and they both miss out on the deal: the end outcome is that they are both left with their original object, which they preferred to swap; rational choices will have led to a sub-optimal outcome.

Some authors disagree with this view and suggest that rationality should lead agents to fulfil the agreement and swap the objects. This objection is based on four arguments. The first argument is that in real world cases agents do not act in a vacuum but are part of a larger society to which moral values apply; breaking such values by cheating may carry an emotional burden or future retribution. I agree with this view but I disregard it in this discussion because it is based on considerations which are external to the game: in the setting I consider there are only two agents carrying out a one-off transaction.

A second argument is based on refusing the paradox: if cheating results in a loss for all agents, there are no grounds to consider it a rational choice.

A third argument is based on the symmetric nature of the game (for an in depth analysis of this topic see (Campbell and Sowden, 1985)); there is a perfect symmetry between the two players: the cost, potential rewards, knowledge, and opportunity for cheating are exactly the same for both; given the situation, if a choice is optimal for an agent, it must be optimal for the other agent too. From this perspective there are only two options for optimality: either they both cheat or they both fulfil the agreement; since cheating results in a sub-optimal outcome for both, the rational choice must be for both agents to adhere to the agreement.

The same conclusion can be reached via an alternative avenue which holds a different meaning in terms of rationality. Because of its symmetry, this problem has a unique optimal solution and two equally rational players will necessarily both seek such a single solution; more important, being rational, they both *know* that the other agent will seek such a solution. This sort of rationality, called super-rationality by Hofstadter (Hofstadter, 1985), will entangle the two agents, making them act as one; as a result, the asymmetric option of a single agent cheating will not just be considered sub-optimal, but will in fact not be available at all.

I subscribe to the view that rationality should lead an agent to fulfil its side of the agreement and propose a further, only slightly different interpretation: because of symmetry, each agent should expect that whatever conclusion it has reached, the other agent will likely have done the same. Consequently, hoping to get a benefit from cheating is based on a gamble: that either the other agent has made a mistake or that it has failed to act rationally; gambling on either option is hardly the hallmark of rationality.

In summary, the prisoners' dilemma offers two views of rational behaviour; one is defined a priori: an agent has to act in order to maximise its individual return in most available situations; this leads to a paradox which, for our discussion, implies that the system does not self-organise optimally. A second view of rational behaviour defies the paradox but is defined a posteriori, only after a careful analysis of the structure of the game; this view leads the system to self-organise optimally. In the next sections I consider how this analysis extends to different scenarios.

3 The tragedy of the commons

The tragedy of the commons (Hardin, 1968) can be seen as a special case of the prisoners' dilemma (Ostrom, 1990): a number of agents has access to a common but limited resource; each agent has an incentive to exploit as much resource as possible; however, if all agents do so, it is likely that the resource will be overexploited and irreversibly crash. In this case, limiting an agent's resource use is analogous to fulfilling an (implicit or explicit) agreement to use the resource sustainably; attempting to maximise an agent's resource use is analogous to cheating.

The discussion on rational behaviour in the previous section appears to extend to this problem: following symmetry considerations, rationality should suggest that each agent limits its resource use and aims for responsible and sustainable management. The argument is valid but not general: it can not be generalised to all settings of a tragedy of the commons. The main difference between the prisoners' dilemma and the tragedy of the commons is that, in the latter, the consequences of cheating may be delayed in time: in real-world versions of the tragedy of the commons, by the time the resource crashes, an alternative resource or source of income may be found; also, discounting future costs and benefits may make it economically valuable to overexploit a resource now rather than conserving it for later use (Walters and Martell, 2004).

It could be argued that accepting the overexploitation of the resource, thereby considering it as a viable option, defies the very meaning of the tragedy of the commons itself and changes the nature of the problem; this is surely correct. Our argument is that there may be conditions under which the viability of the over-exploitation may not be known a priori; while a pre-cautionary principle may suggest not to entertain such option, such a principle itself would need to be defined a priori and it is not explicit in the tragedy formulation. For the purpose of our discussion, it is interesting to notice that the possibility of a change in circumstances does not change the symmetry of a problem but opens the opportunity for the system to self-organise into two states, one of profitable sustainable resource use and one of profitable resource crash; which applies depends on circumstances; the second option was not obvious in the prisoner's dilemma. Unless set a priori, defining rational behaviour in this setting appears problematic.

4 The minority game

Let's now suppose that agents want to exploit a resource distributed in space and can access it at different locations. If the resource is abundant, the agents can decide where to access the resource with few consequences. If the resource is scarce, agents risk to overexploit the resource wherever they access it and the tragedy of the commons described above occurs (Boschetti and Brede, 2009). From a game theoretical and complex system perspective, the interesting scenario occurs when the amount of resource available is comparable with the agents' need. In this setting the location at which agents access the resource affects how much resource can be exploited by the community (Boschetti, 2007), since optimal exploitation happens only when the community spreads its harvesting effort proportionally to the resource distribution. Interesting dynamics occur when no external coordination is available: each agent naturally strives to access the resource at the least exploited location, where it can expect to share the limited resource with the least number of competitors.

This problem has been studied extensively under the name of Minority Game or Bar Problem (Arthur, 1994; Challet and Zhang, 1997; Zhang, 1999). At each iteration, each agent chooses where to access the resource, has the same harvest potential and knowledge of the problem: as before, symmetry applies. According to our previous discussion, rational agents should then realise that the best strategy should be the same for everyone: the most symmetric option is for each agent to choose randomly at each iteration, with probability $\frac{1}{2}$, which of the two areas to harvest.

Interestingly, this approach is not optimal either individually or globally (Savit et al., 1999). For a resource amount smaller than the sum of the total agents' harvest potential the best harvest outcome is obtained when agents act competitively and selfishly by trying to outsmart each other. This is the standard setting of the minority game employed in the literature.

For a resource amount larger than the sum of the total agents' harvest potential the best harvest is provided by the collective intelligence approach (Wolpert et al., 2000). The collective intelligence suggests that individuals should try to maximise the impact their action has on community behaviour. This is 'measured' as the difference in harvest between what the community gains minus what the community *would* have gained *had* that specific agent *not* participated to the harvest. In other words an action which results in an individual gain but no community gain is penalised. For a resource amount larger than the sum of the total agents' harvest potential (but not so large as to make searching for optimal location unnecessary) the collective intelligence allows the agents to spread effort proportionally to the resource distribution. In this setting, the harvest will be maximised (effectively getting very close to optimality) for each agent and also for the overall community (Boschetti, 2007; Brede and De Vries, 2008).

Interestingly for our discussion, this efficient distribution of effort is obtained within a few iterations, after which agents rarely change the harvesting location: very quickly symmetry at the individual level (what location each agent accesses) is broken and symmetry at the community level (even split between locations) is achieved, despite the agents having neither information about the resource distribution nor external coordination.

As a result, an amount of resource roughly equal to the sum of the agents' harvest potential represents a threshold for the community behaviour, which determines whether it is convenient for the agents to choose a collective intelligence approach or a selfish one. In (Boschetti and Brede, 2008) we have proposed a method for dealing with this threshold: agents can adaptively choose whether to employ the collective intelligence or act selfishly by choosing the strategy which provides the largest amount of information about the problem.

What does this discussion imply in term of rational behaviour? In principle, rationality does not necessarily have to match economic performance, for example issues of resilience or alternative non-economic benefits may be accounted for; however, in the problem I discussed these alternative criteria do not apply, since they lie outside the scope of the problem itself. So, as for the prisoner's dilemma, unless we define it a priori, rationality can be evaluated only on issues of symmetry, performance and self-organisation. Still, associating rationality with performance and organisation in the minority game seems to be somewhat unrealistic, since in order to discriminate what the rational behaviour should be, an agent would need to carry out a number of fairly sophisticated steps: first, it would need to record its historical resource exploitation; then, it would have to carry out the computation involved in the collective intelligence approach; then, it would have to evaluate what the 'selfish' return could be; then, it would have to determine which of the previous two provides the largest amount of information about the problem; then it would have to predict at which location it should exploit the resource next; finally it would need to carry out several numerical experiments to convince itself that the approach does indeed work well; this is hardly the intuitive understanding of rationality we are accustomed to. Finding a suitable and general definition of rational behaviour in complex problems is not straightforward.

5 Discussion

In Figure 1 I summarise the analysis of the scenarios in terms of balance between individual and community performance. The X axis represents the performance of an individual agent and the Y axis the community return averaged over the number of individuals. Since the community return is the average of the individual ones, it is impossible for the community return to be larger than the best possible return an individual can obtain: all scenarios have to lie on the right hand side of the 'average global return= best individual return' line (that is, within in the light gray triangular sub-domain).

The best match between individual and community returns is achieved by the Collective Intelligence (Coin in the figure) in the Minority Game (MG in the figure) when the amount of available resources allows this (Boschetti and Brede, 2008). In this scenario individual harvest for both the best and worst performing agents are close to optimal and as a result so is the average community return (Boschetti, 2007). When symmetry is respected in the prisoners' dilemma ('Sym PD' in the Figure) it is impossible for both individual agents to achieve their optimal result (they can't both keep their object and obtain the other agent's one) but they both fulfil their aim by carrying out the swap successfully: both individual and community performances are

good. Stepping down the imaginary ladder of optimality we find the selfish behaviour in the Minority Game ('MG Selfish agent' in the Figure); this is the best the community can do under severe resource constraints (Boschetti and Brede, 2008; Brede and De Vries, 2008). In this case some agents may fare quite well while others may perform quite badly (Boschetti, 2007) and as a result the average community performance is far from optimal. Next we find the asymmetric case of the prisoners' dilemma ('Asym PD' in the figure) in which one agent cheats and the other is cheated; these are represented by two connected ovals in the figure since the occurrence of one implies the other. Here one agent obtains the maximum possible outcome (both objects) while the other obtains nothing at all; the global performance is worse than had they both acted fairly. Finally, the worst global outcome is achieved when both agents cheat in the symmetric prisoners' dilemma.

From the previous discussion it may appear that our analysis of rationality reduces to finding a strategy which provides a maximum return, thereby leading us back to the original definition of economically rational agents we started from. In fact, the previous discussion addressed the question of what implication an individual decision has on the community and thus by self-referentiality on itself. In the optimisation parlance, this is a global, not a local question like the original one.

It is important to remember that the simple games I presented involve dilemmas of an economic nature only. Also, the core of the analysis lies along a line spanning from the individual to the community; in social science parlance, this line connects terms like 'selfish' and 'unethical' to 'generous' and 'morally conscious'. These terms however are not available to this discussion because the mechanistic nature of the games I analysed lacks social context and thus moral implications. In my opinion this strengthens, rather than weakens the argument: even within the vacuum and simplicity of the analysed scenarios a single criterion to define rationality is not available. Lacking social, moral and emotional implications, the culprits for the problem are easy to pinpoint: self-referentiality and the existence of different levels of analysis. The different levels of analysis are where the problem of rationality itself is framed; self-referentiality is what makes any argument bounce endlessly back and forward between the agents and between the levels themselves. It is easy to imagine that the inclusion of social, moral and emotional implications would complicate the analysis further.

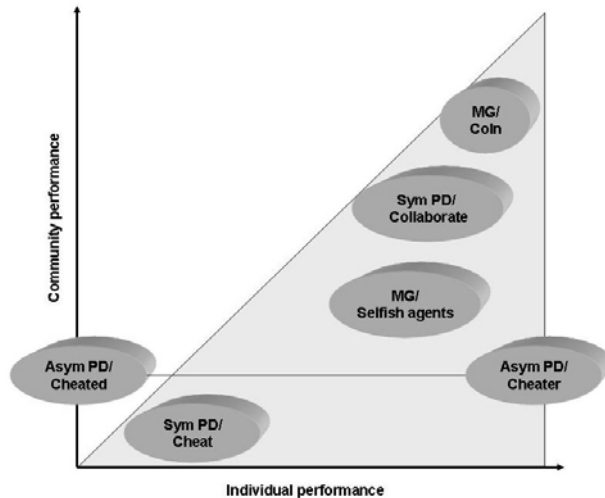


Figure 1. The scenarios discussed are plotted as a function of the individual and community performance they display. MG=Minority Game; Sym PD= symmetric prisoners' dilemma; Asym PD= asymmetric prisoners' dilemma; Coin=Collective Intelligence.

While it is true that self-referentiality and the existence of different levels of analysis are hallmarks of complexity, it is also true that it is hard to imagine any 'simpler' complex problem, that is any complex problem with fewer components, fewer possible scenarios and greater ease of complete definition. As a result, it appears that we are left with three options for defining rational behaviour, each with its benefit and drawbacks:

- 1) we could rely on an a priori definition, for example based on moral or ideological motives (several authors believe the perfectly rational agent of classic economic theory is the result of one such ideology). The benefit of this choice is its simplicity and its intuitive appeal. The drawbacks are logical: first, we may question how such ideology or moral argument arose in the first place; second, it may lead to the paradoxes of rationality discussed above. In the latter case, a rational behaviour may then be seen as a problem to overcome via analysis and experience, which defeats its original intuitive appeal.
- 2) We could work purely a posteriori, first by studying the problem empirically, then by trying to appraise its structure by induction and then by devising a workable, useful and effective definition of rational behaviour. The benefit of this approach is that we could design the definition in such a way that it avoids paradoxes and is 'useful' to the agents in terms of performance and organisation; the drawback lies in its obvious complexity and its likely non-algorithmic nature (Boschetti and Gray, 2007b; a).
- 3) Finally, we could imagine an approach between the previous two: given a problem, a definition of rationality could be used which has proved to be useful in similar settings; this definition could be adopted until a difficulty is encountered, at which point the definition may be modified accordingly. This approach has an evolutionary flavour and it is often suggested that many of our current moral values have evolved over many generations because they provided some sort of group benefit.

6 Conclusions

This analysis cautions against blindly porting values or experiences from one complex problem to another, even those which address something as basic as what it means to be rational. People with extensive real world experience would probably find this unsurprising since many times our judgement is severely tested by complex problems: are the Palestinians or the Israelis right? Should we support the use of genetically modified food in the developing world? That no anchor to rationality is available in moving from something as simple as the prisoner's dilemma to the equally simple minority game may be frustrating, may also confirm that in many cases an empirical, rather than an ideological approach may be needed.

References:

- Arthur, W.B., 1994. Inductive behaviour and bounded rationality. *The American Economic Review*, 84:406-411.
- Axelrod, R., 1984. *The Evolution of Cooperation*. Basic Books, New York.
- Boschetti, F., 2007. Improving resource exploitation via collective intelligence by assessing agents' impact on the community outcome. *Ecological Economics*, 63:553-562.
- Boschetti, F. and Brede, M., 2008. An information-based adaptive strategy for resource exploitation in competitive scenarios. *Technological Forecasting & Social Change* doi:10.1016/j.techfore.2008.05.005
- Boschetti, F. and Brede, M., 2009. Competitive scenarios, community responses and organisational implications. *Emergence: Complexity and Organization*:In print.
- Boschetti, F. and Gray, R., 2007a. Emergence and Computability. *Emergence: Complexity and Organization*, 9:120-130.
- Boschetti, F. and Gray, R., 2007b. A Turing test for Emergence. In: M. Prokopenko (Editor), *Advances in Applied Self-organizing Systems*. Springer-Verlag, London, pp. 349-364.
- Brede, M. and De Vries, H., 2008. Harvesting Heterogeneous Renewable Resources: Uncoordinated, Selfish, Team-, and Community-Oriented Strategies. *Environmental Modelling & Software*:Submitted.
- Campbell, R. and Sowden, L., 1985. *Paradoxes of rationality and cooperation : prisoner's dilemma and Newcomb's problem*. University of British Columbia Press, Vancouver.
- Challet, D. and Zhang, Y.C., 1997. Emergence of cooperation and organization in an evolutionary game. *Physica A*, 246:407-418.
- Hardin, G., 1968. The tragedy of the commons. *Science*, 162:1243-1248.
- Hofstadter, D., 1985. *Metamagical Themas*. Basic Books, New York.
- Ostrom, E., 1990. *Governing the Commons: The Evolution of Institutions for Collective Action* Cambridge University Press, Cambridge.
- Savit, R., Manuca, R. and Riolo, R., 1999. Adaptive Competition, Market Efficiency, and Phase Transitions. *Physical Review Letters*, 82:2203-2206.
- Walters, C. and Martell, S., 2004. *Fisheries ecology and management* Princeton University Press.

Wolpert, D., Wheeler, K. and Tumer, K., 2000. Intelligence for control of distributed dynamical systems. *Europhysics Letters*, 49.

Zhang, Y.C., 1999. Modeling market mechanism with evolutionary games. *Europhysics News*, 29:51-53.