Chapter 1

Introduction to Collective Systems

A collective system is a large system of adaptive agents, where each agent has her own utility function to optimize, along with global performance measures of the full system. The envisioned objective is to study the mechanism of inducing desirable collective outcomes. This aim is quite novel, since a collective of agents needs to establish coordinated and synchronized behavior from the bottom up. In this chapter, we provide a survey of approaches to the study of collective systems.

1.1 Collective Outcomes of Interacting Agents

Billions of people make billions of decisions everyday about many things. It often appears that the aggregation of these unmanaged individual decisions leads to a desired outcome. It is amazing that economic and social activities generally work well in this way without any authority. Adam Smith characterized this fact by stating that an "*unseen hand*" brought about coordination among self-interested individual economic activities. The unseen hand is observed behind many market activities. This principle also works as a basic mechanism for allocating limited resources to people who need them.

People constantly interact with each other in different ways and for different purposes. Somehow these individual interactions produce some coherence at the aggregate level, and therefore, aggregation may generate structure and regularity. The individuals involved may have a very limited view of some part of the whole system but their activities are coordinated to a large degree and produce a desirable outcome at the aggregate level.

However, there are other systems for which it is difficult to understand how they work or to find out better ways to make them work. For instance, many economic and social systems often produce inefficient outcomes at the aggregate level in a way that the individuals who comprise the system need not know anything about or even be aware of it. When the system results in some undesirable outcome, we often think about whether it is due to the members who comprise the system. We tend to observe the resulting outcome as corresponding to the intentions of the members who compromise the system.

There is strong interest in many fields to answer the following questions. How do interacting individuals with micro-motives produce the aggregate outcome? How do we identify the micro-rules of agents that produce some regularities of interest at the macroscopic level? There has been no natural methodology for systematically studying these issues.

Most of our social activities are substantially free of centralized management, and although we may care how it all comes out in the aggregate, our own decisions and behaviors are typically motivated by self-interest. Therefore, in examining collective behavior, we shall draw heavily on the individual decisions. It might be argued that understanding how individuals make decisions is sufficient to understand most parts of the collective system. Although individual decisions are important to understand, they are not sufficient to describe how a collection of agents arrives at specific decisions. These situations, in which the decision of an agent depends on the decisions of others, are situations that usually do not permit any simple summation or extrapolation to the aggregate (Schelling,1978). To make this connection, we usually have to look at the system of interactions among agents.

We usually ascribe human behaviors as if they are oriented toward a goal. Peoples have preferences and pursue their own goals, or maximize comfort as well as minimize effort embarrassment. We might characterize these behaviors as *purposive behaviors*. Economists argue that much of individual private consumption is also dependent upon other peoples' consumption. We often behave by reacting to others. Therefore, what we also have is a mode of *contingent behavior* that depends on what other people are doing (Schelling, 1978). For example, each person's enjoyment of driving a car is inversely related to others' enjoyment if too many peoples drive. Everybody becomes stuck in congested traffic in this case. This is a kind of social congestion and the problem is that there is no way of knowing what others will do. When we are in a mode of contingent behavior, the resulting collective behavior is often volatile and far from desirable.

It is not easy to tell from collective phenomena just what the motives are behind individuals and how strong they are. For instance, consider a traffic jam again. It is not easy to capture the properties of a traffic jam at the aggregate level without describing what individual drivers do. Each of these drivers is different, and the characteristics of their driving behavior become the rules in the model. When we run this model we can reproduce a traffic jam, but this time we need to watch closely how the individual drivers interact with each other and we can inject to see how these interactive behaviors among drivers would affect the visible properties of the traffic jam (Resnick, 1999)(Bonabeau, 2002).

Therefore, we have to look closely at agents who are adapting to other agents. In this way, the behavior of one agent affects the behaviors of the other agents. How well agents accomplish what they want to accomplish depends on what other agents are doing. What makes this kind of interactive situation interesting and difficult is that the aggregate outcome is what has to be evaluated, not merely how agents behave within the constraints of their own environments.

How well they do for themselves in adapting to environments is not equivalent to how satisfactory a social environment they collectively create for themselves. There is no presumption that the self-serving behavior of agents should lead to collectively satisfactory results. If our problem is that there is too much traffic, we are also part of the problem. If we raise our voice to make ourselves heard, we add to the noise level that other people are raising their voices to be heard over.

Our complex systems often result in the features of *emergent properties*, which are properties of the system that separate components do not have. These emergent properties, we find, are the result of not only the behavior of individual agents but the interactions between them as well. For instance, what drivers do on the road depends on what other drivers do. This can not be explained without looking at how the agents behave and interact with each other to make up the whole. Resulting traffic jams are counterintuitive phenomenon that we could only predict with the framework of the collective system of interacting agents.

We can observe many collective phenomena viewed as *emergence* that has arisen from billions of small-scale and short-term decisions of interacting agents. Viewing complex systems as a collective of interacting agents means adopting a new scientific approach that shifts from *reductionism* to *connectionism.*

With the view of reductionism, every phenomenon we observe can be reduced to a collection of components, the movement of which is governed by the deterministic laws of nature. In such reductionism, there seems to be no place for novelty or emergence. The basic approach with the view of reductionism is the rational choice model. The rational choice theory posits that an agent behaves to optimize her own utility produces relevant and plausible prediction about many aggregate phenomena.

However, there are many critics of approaches based on the *rational choice model.* The problem of the rational choice model is that it assumes agents who are sufficiently rational. Goals and purposes of agents are also often related directly to other agents or they are constrained by an environment that consists of other agents who are pursuing their own goals.

When a society or organization faces some complex problems, the typical reaction is to fall into "*centralized thinking*" (Watts, 2001). A small coherent group of experts decide what to do based on the characteristics of the problem, and execute rules and everyone else then simply follows these rules. However, introducing additional rules can serve only to make the problem worse. This is because it is usually centralized thinking behind these local rules, so the effect of new local rules being added to existing local rules is quite strong. Without modeling the process of the chains of reactions, it would be very hard for a human brain to predict this pathological collective behavior.

To understand this paradox, we need to take a look at the problem of "*decentralized thinking*" (Resnick, 1999). What should be clear is that combining the many different individuals involved at a single point is almost certain not to succeed in delivering the kind of essential functionality. Some other kind of connectionism is required.

1.2 The Study of Collective Systems

A collective system is modeled as a collection of autonomous decision-making entities, called agents. In this section, we provide the definition and a survey of approaches to collective systems. A collective system, or just simply a collective, means any complex system of interacting agents, together with performance measures by which we can rank the behavior of the entire system (Tumer and Wolpert, 2004). Collective systems include a collection of diverse and mutually adaptive agents pursuing varied and often conflicting self-interests.

Many organisms form aggregations that have strong effects on individual behaviors. Familiar examples include schools of fish and flocks of birds. Auyang (1998) defines the term "*collective*" for such aggregations. According to Auyang, the defining characteristics of a collective are as follows. Interactions among individuals making up a collective are strong, that is internal cohesion is strong while external interactions are weak. Furthermore, collectives have their own characteristics and processes that can be understood independent of the individuals that compromise them.

Another defining characteristic of collectives in ecological systems is that collectives exist for longer or shorter times than do the individuals making up the collective. Collectives can be treated as an additional level of organization between the individual and the population (Grim, 2005). Individuals belonging to a collective may behave very differently from individuals alone, so different traits may be needed to model in individuals that are not in a collective.

The behavior of a collective emerges from traits of individuals. A school of fish is an example of modeling a collective as emerging from relatively simple traits of individuals, and these traits give rise to individual behaviors that form the collective. Representing a collective explicitly does not mean that individuals are ignored. Instead, a collective can also be represented by the manner in which individual behaviors affect the collective and how the state of a collective affects individual behaviors. For instance, individuals make decisions as to when to disperse, and this affects the formation and persistence of the collective, while these individual decisions are based in part on the state of the collective. Therefore, the collective system can only be understood by modeling individuals and the aggregate, as well as the link between them.

We use the term *collective system* when it is impossible to reduce the overall behavior of the system to a set of properties characterizing the individual agents. Interaction between agents is also important consideration that produces emergent properties at the aggregate level that are simply not present when the components are considered individually. Another important feature of collective systems is their sensitivity to even small perturbations. The same action is found to lead to a very broad range of responses, making it exceedingly difficult to perform prediction or to develop any type of experience of a typical scenario.

At the basic level, a collective system consists of agents and the relationships between them. Agents may execute various behaviors that are appropriate for the system they represent. Each agent individually assesses her situation and behaves on the basis of a set of local and idiosyncratic rules. Repetitive and competitive interactions between agents are a feature of a collective system and require the power of computers to explore the collective dynamics, which are not obtainable through pure mathematical modeling and analysis.

An agent is described by a number of fundamental components including, (1) a private utility function, (2) the drive to optimize the private utility function, (3) a set of possible actions, (4) the rule that generates an action in an attempt to optimize her utility or adapt to others, and (5) information and memory about history. In addition, (6) agents may be capable of learning or evolving, allowing new behavioral rules to emerge.

The collective systems are situated between a few agents in game theoretic systems, a few hundred agents in multi-agents systems, and a larger scale of agents in typical economic and social systems. Hardware development will soon make possible the construction of very large-scale models. Therefore, we may obviate the need for small-scale multi-agent systems. It will be argued that the main impediment to creating empirically relevant agents on this scale is our current lack of understanding of the realistic behavior of agents. Therefore, the bottleneck of what rules to decide for agents is also the primary challenge for researching collective systems (Axtell and Epstein, 1999).

The performance of the collective system, which consists of many interacting agents, should be described on two different levels: the microscopic level, where the decisions of the individual agents occur, and the macroscopic level, where the collective behavior can be observed. Understanding the role of the link between these two levels also remains a challenge (Schweitzer, 2002).

There are two related theoretical issues in the study on collective systems. One is the effect of interactions among agents in determining macroscopic outcomes. The other issue is how to design the micro-rules of agents that produce a desirable outcome at the macroscopic level.

Tumer and Wolpert (2004) propose two different perspectives for the study of collective systems: analysis, or the *forward problem*, and design, or the *inverse problem*.

(1) The forward problem focuses on how the localized attributes of a collective (the properties of agents) induce global behavior of interest and determine system performance.

(2) The inverse problem arises when we wish to design a system to induce a desirable outcome.

It is generally believed that ants in an ant colony do not know exactly how to build the ant colony in which they live. Each ant has certain things that it does, in coordinated association with other ants, but there is no ant that designs the whole colony. No individual ant knows whether there are too few or too many ants exploring for food. Why the colony of ants works collectively as it does, and as effectively as it does, remains a mystery. An important factor in understanding such behavior is the interactions among ants (Bonabeau, 1999).

Economists may not like the idea of comparing the economy to an ant colony. They are no doubt convinced that such organizations are in some sense optimal, but they are not convinced that the optimality is achieved in the same way as it is in a market. Thus, if we ask most economists to describe the basic question that concerns them, they answer that they are trying to understand the equilibrium of the market and whether it entails an efficient use of limited resources (Kirman, 2001).

Most of these are activities in which an agent's behavior is influenced by others, or in which agents care about the behavior of others. Or, they both care and are influenced by trying to obtain an equilibrium. An equilibrium is a stable situation in which some motion, or activity, or some adjustment, or response, has ceased, resulting in some stagnation in which several items that have been interacting and adjusting to each other are at last adjusted and are *in balance*.

Equilibrium is a central concept in the study of social systems as well as in the study of physical collective phenomena. In physical systems, equilibrium results from a balancing of forces. In a physical system, particles are in equilibrium when they do not deviate from a given position or stable trajectory. In a collective of agents, their behaviors are typically motivated toward their own interests. Therefore, in a collective of agents, equilibrium is a balancing of intentions. That is, individuals' intentions are in equilibrium when no one wants to deviate from her intended behavior given the intentions of others (Young, 2005).

However, it is widely observed that an equilibrium situation is not usually efficient at the macro level. The fact that selfish behavior at equilibrium may not achieve full efficiency is well documented in the literature. While all agents understand that the outcome is inefficient, acting independently cannot manage the collective with respect to what actions to take or how to decide these actions.

Yet there is also the problem of explaining how disparate individual activities are coordinated, as in the case of ants and other social insects. The solution must apply equally well to other collective systems. The forward problem arises in the study of some previously existing field such as complex theory. The fundamental problem lies in determining how the combined actions of many agents lead to coordinated behavior on a global scale. Approaches in existing research fields may provide valuable insight into some aspects of studying the forward problem of collective systems. However, these approaches fall short of providing suitable methodologies for dealing with heterogeneity in agents and their interactions.

Agent interactions also occur on structured networks. Yet the development of tools for modeling, understanding, and predicting dynamic agent interactions and behavior on complex networks lags behind. Even recent progress in complex network modeling has not yet offered any capability to model dynamic processes among agents who interact on a global scale, as in small-world or scale-free networks. Computational modeling of dynamic agent interactions on structured networks is important for understanding the sometimes counter-intuitive dynamics of such loosely coupled agent systems of strategic interactions.

Given a collective system, there is also an associated *inverse problem* or a *design problem*, with respect to configuring or modifying each agent so that in the pursuit of their own interests, the agents also optimize the global performance. Solving this kind of inverse problem may involve determining and modifying the internal models of agents as well as the method of interactions between agents.

We should also consider the degree of freedom that each agent should have. Here, there are two basic approaches: top-down and bottom-up approaches. With the top-down approach, the designer may have the freedom to assign the private utility functions of the agents. With the bottom-up approach, agents may have the freedom and incentive to modify their private utility functions. In either case, the focus is on guiding collective systems toward desirable outcomes.

The agent wishes to optimize her own utility, and the system designer wishes to implement a decentralized algorithm for optimizing the collective performance. However, excessive pollution or social congestion problems are the most commonly recognized examples of a break between individual optimization and collective efficiency. Therefore, the inverse problem is concerned with how the private utility functions of agents can be redesigned so that their selfish behaviors give rise to a desired collective outcome. If the collective system can be designed, we need design it so that selfish behaviors of agents need not degrade the system performance, by providing a carefully chosen decentralized mechanism that can be implemented at the agent level.

The inverse problem is a high priority for the study of collective systems. The investigation of the inverse design problem also comes up in the study of already existing fields, such as learning in games and multi-agent systems. However, these approaches are basically based on so-called equilibrium analysis and, therefore, they fall short of providing suitable methodologies for solving the inverse problem.

1.3 The Price of Anarchy in Collective Systems

The fact that selfish behavior may not achieve full efficiency has been well known in the literature. Therefore, it is important to investigate the loss of collective welfare due to selfish and uncoordinated behavior. Recent research efforts have focused on quantifying this loss for specific game environments. The resulting degree of efficiency loss is known as the *price of anarchy* (Roughgarden, 2005). The investigation of the price of anarchy provides a foundation for the design of collective systems with robustness against selfish behaviors. We may need to design systems such that selfish behaviors of individuals need not degrade the system performance.

Social interactions in which people's behavior is influenced by the behaviors of others, or in which people care about the behaviors of others, are analyzed in the context of *collective action*. In this section, I demonstrate that there are a host of collective action problems that share the same general structure and that make agent interaction problematic whenever they arise.

An *externality* is an unexpected side-effect of the social activities of some individuals on seemingly unrelated people. An externality also occurs when individuals care about the choices of others, and each individual's choice affects the choices of others. There are basically two types of social activities with externalities: *strategic compatibility* and *strategic complementarity*. With strategic compatibility, individual payoffs increase with the number of people taking the same action. Instead, with strategic complementarity, payoffs are better if the actions of people are distributed.

Many collective action problems can be investigated through the underlying *social games*. Game theory studies the interaction between human beings and provides a way of modeling strategic interactions, situations in which the consequences of agents' actions depend on the actions taken by the others, and each agent knows who is involved in the same game. The outcome of a conventional game is a set of actions taken by all involved agents. On the other hand, social games are a way of modeling social interactions in which each agent may not know who is involved.

1.3.1 Collection action with strategic compatibility

The study of collective action began with Olson (1965). In his book, there are rich discussions about how different factors, such as instrumental and social incentives, are embodied as the payoff structures in a collective action problem. Consider the provision of a *public good* for individuals. In contrast to private goods, public goods are non-excludable in consumption, and the nature of the public enables an individual to have a free ride. Olson analyzed this problem in terms of the size of the group for which the public good is provided. Olson found that unless the number of individuals in the group is quite small, or unless there is coercion or some other special device to make individuals act in their common interest, self-interested individuals will not act to achieve collective interest, and simply attempt to gain a free ride.

Free riding is a very frequent phenomenon in everyday life. Economists use public goods games to calculate optimal taxes and subsidies. These exercises rest on the assumption that agents will free ride on others, hence leading to a social inefficiency. An effective approach to handle the problem of free riding is to use external enforcement. However, an alternative approach has emerged that investigates the exact circumstances under which efficient collective outcome is possible in the absence of external enforcement.

However, rather than exploring the possibility that various types of games coexist, most collective action problems focus on analyzing cases in which the collective action can be interpreted as a type of dilemma game. In general, human interaction is characterized by a mixture of conflict and consensus.

Let us consider situations in which N agents are identically situated by presenting a binary choice problem with externalities. That is, each agent's payoff, whichever way she makes her choice, depends on the number of agents who choose one way or the other. The typical situation is the interaction with strategic compatibility, where the increased effort by some agents leads the remaining agents to follow suit, which produces *multiplier effects*.

We formally consider this situation by specifying the payoff function of each agent with the following two strategies:

$$\begin{aligned} &S_1 : \text{Disclose her private knowledge,} \\ &S_2 : \text{Does not disclose.} \end{aligned} \tag{1.1}$$

Each agent receives a benefit in proportion to the number of agents to disclose by choosing S_1, which is denoted by n $(0 \le n \le N)$. The payoff function of each agent is defined as follows:

$$\begin{aligned} &U(S_1) = a(n/N) - c\,, \\ &U(S_2) = b(n/N)\,. \end{aligned} \tag{1.2}$$

If an agent discloses her knowledge, she receives some proportional benefit $a(n/N)$ minus some cost c due to disclosure. Even if she does not disclose her knowledge, she can receive some benefit $b(n/N)$ from the contribution of other agents as the *spillover effect*.

Interdependent decision making problems involving N agents are called *N-person games*. We simplify an *N-person game* considering a population of N agents, each with a binary choice between S_1 and S_2. In addition, for any agent the payoff for choosing either S_1 or S_2 depends on how many other agents choose S_1 or S_2. This social game can be analyzed in the following two cases.

<Case 1> $b > a - c$: In this case, the payoff to S_2 is greater than that to S_1. Therefore, the rational choice of each agent is S_2 without considering the choices of the other agents. In game theory, this is defined as a dominant strategy. When this condition holds for all agents, no agent will trade, and this results in the sharing of no common knowledge. This case is known as the *N-person prisoner's dilemma game* (NPD) or the *social dilemma*.

It may not be surprising that the result of local optimization by many agents with conflicting interests does not possess any type of global optimality. In the language of game theory, this means that an equilibrium situation arising from individual rationality can be *Pareto-inefficient*, and thus the outcome can be more efficient, that is, some are better off while no one is worse off. In game theory, the most canonical example of Pareto inefficiency of selfish behavior is the social dilemma. There are many problems involving the clash of individual and collective interests, including the energy crisis problem, various problems related to the conservation of scare natural resources, and a range of problems arising from environmental pollution.

Game theory suggests two alternative solutions to social dilemmas. One solution is to introduce external enforcement. In this case, the payoff structure is altered in such a way that the defecting person incurs some penalty. The other solution is to repeat the game in a way that, from the standpoint of the players, looks like it is being played infinitely many times. The evolutionary paradigm is also built on this type of analysis. It should be noted that the analysis of repeated games is also a very fertile area of study with respect to collective systems.

<Case 2> $b < a - c$: It is easy to recognize that there are other collective action problems with a very different structure from that of the social dilemma. Under the condition of Case 2, we have two stable solutions: an all-S_1 choice and an all-S_2 choice.

The payoff is maximum at $n/N=1$, when all agents choose S_1, and they can enjoy the highest externality, which is better for all agents. On the other hand, if the proportion of agents who disclose their knowledge is relatively low (less than $c/(a-b)$), it becomes rational to choose S_2. If this condition holds for all agents, no agent will disclose her knowledge, and they encounter the same situation as the social dilemma.

In this case, with multiple equilibria, the problem involves not only how to get a concerted choice, but also how to achieve the best equilibrium. If many agents choose S_2, no agent is motivated to choose the inferior choice S_2 unless a sufficient number of agents switch beyond the intersection of the two payoff functions in (1.2). Therefore, the ratio at the intersection provides a crucial mass parameter for the selection of the efficient equilibrium.

It is enough merely to get agents to make the right choice at the beginning. If the ratio of agents that choose the superior strategy (S_1) is greater than that of the intersection point of $U(S_1)$ and $U(S_2)$, then all agents will self-enforce to choose S_1. In this sense, a certain threshold appears. If the initial ratio of the agents choosing S_1 exceeds this threshold, they can induce other agents to shift to a superior choice. The inverse is also true. If many agents stick to an inefficient choice, then all agents follow the same path to an undesirable outcome.

1.3.2 Collection action with strategic complementarity

The fact that selfish behavior need not produce a socially optimal outcome was well known before the advent of game theory. Pigou (1920) proposed a route selection problem in which individuals independently need to travel from the same source to destination. Suppose that there are two highways between two locations. One of which is broad enough to accommodate all traffic that appears without congestion, but is poorly graded and surfaced. While the other is a much better road, but is narrow and quite limited in capacity. Assuming that all individuals aim to minimize the driving time, we have good reason to expect all traffic to follow the better road, and therefore, the better road will be completely congested.

The route selection problem is commonly used for the prediction of traffic patterns in transportation networks that are subject to congestion. We formulate the route selection problem as follows. There are two alternative choices, *Route A*, using a private vehicle, or *Route B*, using a public train to commute to the same destination. Let us suppose that the required time if an agent chooses public transportation, the train (Route

B), is 40 minutes, which is constant regardless of the number of agents on the train. On the other hand, the required time for an agent who chooses a personal vehicle (Route A) is an increasing function of the number of agents who choose the same route, as depicted in Figure 1.1 (Dixit and Nalebuff, 1991). If a large number of agents are free to choose either of the two choices, they will tend to distribute themselves between the two routes in such proportions that the transportation time will be the same for every agent on both routes. As more agents use personal vehicles (Route A), congestion develops, until at a certain point there is no difference between routes.

Knight (1924) developed the idea of traffic equilibrium. He gave a simple and intuitive description of a postulate of the route choice under congested conditions. Wardrop (1952) clarified two basic principles that formalize the notion of *user equilibrium* and *system optimal*. The latter is introduced as an alternative behavior postulate of the minimization of the total travel costs.

Wardrop's first principle is that the journey times in using two routes are equal and are less than those that would be experienced by a single vehicle on any unused route. Each user non-cooperatively seeks to

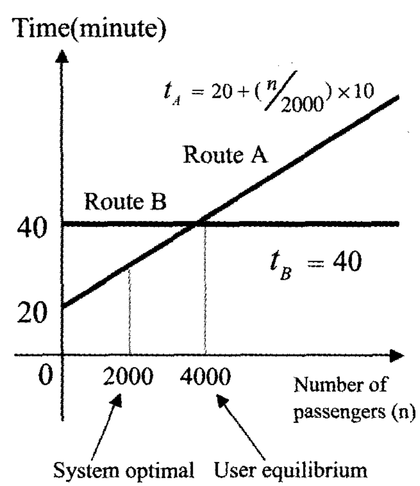

Figure 1.1 The vertical axis represents the estimated traveling time, and the horizontal axis represents the number of agents that use private vehicles (Route A)

minimize her cost of transportation. This principle of route choice, which is identical to the notion postulated by Knight, became accepted as a sound and simple behavioral principle to describe the spreading of trips over alternate routes due to congested conditions. The traffic flows that satisfy this principle are usually referred to as *user equilibrium*, since each user chooses the route that is the best. Specifically, a user-optimized equilibrium is reached when no user can lower her transportation cost through unilateral action.

On the other hand, the system optimal conditional is characterized by Wardrop's second principle, which is the equilibrium at which the average journey time is minimum. This implies that each user behaves cooperatively in choosing her own route to ensure the most efficient use of the entire system. Traffic flows satisfying Wardrop's second principle are generally known as *system optimal*. This second principle may require that users cooperate fully or that a central authority controls the transportation system.

User equilibrium is realized at the intersection in Figure 1.1. However, the system optimal condition is achieved at $n=2,000$, which is half of user equilibrium. In much of the transportation science literature, *Nash equilibrium* is called *user equilibrium*, and *Pareto-efficient outcome* is called system optimality.

Can the power of centralized control improve over the selfish outcome? We suppose that we can choose who chooses what, and assigning half of the commuters to each of the two alternative choices. The commuters who are assigned to the train are no worse off than in the previous outcome. On the other hand, the commuters that are allowed to drive now enjoy lighter road traffic conditions and arrive at their destination at half of the previous time. Therefore, the state of affairs has improved for half of the commuters and has not been changed for the rest of the commuters. The problem of equity or fairness may arise with regard to who should be better off.

A solution to this kind of problem may invoke the intervention of an authority that finds the system optimal condition and imposes the desired behavior on agents. While such an optimal solution may be easy to find, the implementation of a desired solution becomes difficult to enforce in many practical situations.

1.4 Review of the Literature

The emerging research field of collective systems represents a fundamentally new scientific approach to sharing compared to existing approaches. Collective systems are situated between a few agents in game-theoretic systems, among a few hundred agents in multi-agents systems, and among a greater number of agents in typical social systems. Some relevant research fields provide a partial solution to the study and design of collective systems. In particular, game theory, multi-agent systems, complex systems, and evolutionary systems are fields that grapple with some of the issues encountered in the field of collective systems. What is needed for exploring new research fields is to survey the basic concepts in related fields and to clarify where they fall short of providing suitable methodologies for the study of collective systems.

<Agent-based modeling> The approach of agent-based modeling is the main tool in many research fields. With agent-based modeling, we can describe a system from the bottom up, from the point of view of its constituent units, as opposed to a top-down approach, where we look at the properties at the aggregate level without worrying about the system's components and their interactions. The novelty in agent-based modeling, compared to what physicists call micro-simulation, is that we are dealing with modeling collective systems, where the components of the system are agents or human beings with adaptive and evolving behavior.

It is generally not possible to conclude for an agent-based model that a particular attribute will give an agent an absolute advantage over time, or that a particular behavioral rule is optimally configured for an agent in an absolute sense (Epstein and Axtell 1996). In principle, using agent-based tools, a modeler can permit attributes and rules to vary and evolve over time. These variations could be the result of innate or external forces for change, or they could result from deliberate actions undertaken by agents. We also relax assumptions to permit endogenous adaptation or learning. This raises an interesting nature-nurture modeling issue: namely, which attributes and rules of agents should be viewed as part of their core maintained identities and which attributes and rules should be permitted to vary in response to environmental influences? Moreover, this issue arises at both the individual and collective levels. How much variation in behavioral rule should agent to be permitted to exhibit over time, and how much variation should be permitted across all agents?

<Multi-agent systems> Today we face many challenges with respect to designing large-scale systems consisting of multiple autonomous agents. Research on multi-agent systems is intended to provide both principles for construction of complex systems involving multiple agents and mechanisms for the coordination of many agents. The most important reason is that some domains require multiple agents. In particular, if there are different people with different goals and proprietary information, then a multi-agent system is needed to handle their interactions.

An agent-based model is extended to the study of multiple agents and their interactions. We specify how multiple agents interact, and then observe the properties that occur at the macro level. The connection between micro-motivation and macro-outcomes will be developed in which agents are instantiated to interact according to fixed or evolving local rules.

An important characteristic to consider when designing a multi-agent system is whether the agents are stable or evolving. Because of the inherent complexity of the problem domain, there is a great deal of interest in using machine learning techniques to help handle this complexity. Of course agents with learning capabilities can be particularly useful in dynamic environments, but when agents are allowed to learn or evolve into other agents, it is not guaranteed that the outcome will be desirable.

Agents are both heterogeneous and versatile. As a result of both behavioral heterogeneity and versatility, small differences in agents can make large differences in collective outcomes. Heterogeneity turns up repeatedly as a crucial factor in collective systems. But the situation is not always as simple as saying that heterogeneity is desirable and homogeneity is not good. The basic question remains: what is the proper balance between heterogeneity and homogeneity? When heterogeneity is significant, we need to be able to show the gains from heterogeneity.

However, the analysis of a collective system of heterogeneous agents becomes difficult and is often intractable.

Another interesting problem is that agents are very homogeneous in the beginning. Differences in behavior and strategy use evolve endogenously as the collective system runs. Agent heterogeneity becomes a changing feature of the collective systems that can then be studied. Unlike some approaches in the previous research, we are primarily interested in the problem in which the preferences, and even the identities of the agents, can evolve over time, rather than situations in which the agents and their preferences are fixed.

<Complex systems> Complex systems deal with systems that are composed of many interacting particles. Complex systems often result in features, self-organization and emergent properties, which are properties of the system that the separate parts do not have. Therefore, the emerging system outcome is extremely hard to predict.

Axelrod and Cohen (2001) propose three methods by which to harness the complexity based on variation, interaction and selection. The term *harnessing complexity* means deliberately changing the structure of a system in order to increase some measure of performance. Variation in a population of agents and actions of agents provides the raw material for adaptation. However, we need to select the proper balance between variety and uniformity. The mechanism that deals with interactions fits into two types: external and internal. The external mechanism is a way to modify the system from outside. On the other hand, the internal mechanism is a way to change the interaction patterns that are driven by the components of the system.

Selection is based on natural selection in evolutionary biology. The selection mechanism is important as the fundamental means by which agents and actions should be eliminated and replaced. While natural selection provides an important paradigm for how an evolving system can work, it also has a serious disadvantage compared with collectives, where we are interested in achieving desired adaptation and evolution. Furthermore, there are two approaches to selection: selecting at the agent level and selecting at the strategy level. Selection at these two levels can work very differently. Determining whether selection is at the agent level or strategy level depends on the performance measures. Using fine-grained and short-term measures of success can help individual learning by providing focused and rapid feedback. Such narrow and prompt measures of success can also be used to evaluate who is successful and who is not.

However, our challenge in collective systems in dealing with the overall performance or long-term measures, using fine-grained measures of success (individual utilities) can easily be misleading. In collective systems, individual measures need to be appropriately correlated, so that agents can generally use strategies that are mutually beneficial. The importance of interaction may not be understood if selection is done at the agent level.

<Complex adaptive systems> How does self-organization work? This is a huge question that humans may never answer completely. Evidently something intrinsic in the manifestation of the reality that humans perceive may spontaneously produce spatial, temporal, or functional structures by means of self-organization, the principles of which are searched in a variety of disciplines, ranging from physics, chemistry and biology, to medicine, psychology and sociology.

Complex adaptive systems proposed by Holland (1995) mainly concern self-organization and emergence in complex large-scale behaviors from the aggregate interactions of less complex agents. An ant nest serves as a typical example. The individual ant has a stereotyped behavior, and almost always dies when circumstances do not fit the stereotype. On the other hand, the collective of ants, the ant nest, is adaptive, surviving over long periods in the face of a wide range of hazards. It is much like an intelligent organism constructed of relatively unintelligent parts (Bonabeau, 1999).

The process of self-organization can be accelerated and deepened by increasing variation, for example by adding noise to systems. Collective systems may have several equilibrium outcomes. To adapt to a changing environment, the systems need a sufficiently large variety of possible equilibrium states to cope with likely perturbations. Given this variety, the most adequate configurations are selected according to the fitness defined for the system. The basic method is then to define an appropriate fitness function that distinguishes better outcomes from worse outcomes, and then create a system in which the components (agents) vary relative to each other in such a way as to discover behavioral rules with higher fitness. This is also a challenging issue in the study of collective systems.

<*Evolution*> Evolution is based on the concept of natural selection that supports the survival of more successful strategies or individuals. In general, an evolutionary process combines two basic elements: A *mutation* mechanism that provides variation and a *selection* mechanism that favors some variations over others. Agents with higher payoff are at a productive advantage compared to agents who use strategy with lower payoff. Hence, the latter decrease in frequency in the population over time by natural selection.

It is not surprising that many scientists are exploring a new unified theory of evolution by merging learning in game theory and evolutionary game theory with modern biological evolution theory. This new theory attempts to explain all kinds of evolutionary processes. Its methods and models in fact cover not only biological evolution of organisms but also the evolution of animal and human behavior in their societies.

There may be two competing approaches in dealing with evolution in collective systems: the microscopic model based on individual learning and the macroscopic model based on system evolution. Instead of drawing a distinction between models of learning and evolution, it is easier to make a distinction between models that describe adaptive behavior at the individual level and those that describe adaptive behavior at the aggregate level.

<*Co-evolution*> If we observe the natural world, species do not evolve in isolation, but rather, they have, to varying degrees, an evolutionary history of interactions with other species. Much of the diversity and specialization observable within the natural world is due to *co-evolution*.

Like biological systems, many socio-economic systems can be modeled as a co-evolutionary system containing many interactive agents, each reciprocally evolving in response to adaptations in the others. If multiple populations of agents are adapting to each other, then the result is a co-evolutionary process. If we use the standard definition of co-evolution, there must be two populations, with each reciprocally evolving specific adaptations and counter-adaptations in response to the other. Co-evolution is defined as an evolutionary change in a trait of the individuals in one population in response to a trait of the individuals of a second population, followed by an evolutionary response by the second population to the change in the first.

Co-evolution is a holistic, synergetic and complex evolutionary flow that cannot be split up into components. Co-evolution rests not only on mutually coupled interactions, but also on our desire to realize better outcomes by solving mutual conflicts or overcoming competition. However, the problem to contend with in co-evolution is the possibility of an escalating *arms race* with no end. Competing agents might continually adapt to each other in more and more specialized ways, never stabilizing at a desirable behavior. This is an example of the problem of sub-optimization. Optimizing by each individual does not lead to optimal performance for the collective system as a whole.

<Evolutionary dynamics > The term *evolutionary dynamics* often refers to systems that exhibit a time evolution in which the character of the dynamics may change due to internal mechanisms. Such models are of course interesting for studying systems in which variation and selection are important components. Evolutionary dynamics are described by equations of motion that may change in time according to certain rules that can be interpreted as mutation operations.

For a species, survival is a necessary goal in any given environment. On the other hand, for a collective system, both purpose and environment need to be specified by the designers or agents who compromise the system. If certain aspects of the world can be set by design, one can explore through intensive experimentation, in which designs tend to induce desirable outcomes when other aspects of the world are permitted to exhibit realistic degrees of plasticity. Alternatively, exploiting the growing power of evolutionary algorithms, one can deliberately induce evolution as a means of discovering improved design configurations.

One important area of research on collective systems lies outside the conventional evolutionary approach based on the Darwinian paradigm of natural selection. Co-evolution also concerns cooperation within in and between species. For instance, in symbiosis, competition is suppressed because the long-term benefits gained from cooperation outweigh short-term competitive advantages. A mathematical framework to model co-evolutionary dynamics in such non-Darwinian systems has been developed (Crutchfield and Schuster, 2003).

<Individual optimization and collective efficiency> We might expect collective behavior to be closer to the optimal behavior if we typically assume rational behavior at the individual level. Adam Smith's conclusion that collective efficiency arises from the individual pursuit of self-interest may be more general than it appears. The connection between individual rationality and collective efficiency, between optimization by individuals and optimality in the aggregate, has been studied in some domains. Regarding this issue, the traditional approaches usually assume that aggregate efficiency requires individual optimization.

Collective behavior may be rational, whereas that of the individuals may not be so. Gode and Sunder (1993) show that *market efficiency*, a key characteristic of market outcomes, is largely independent of variations in individual behavior under classical condition. They showed that market efficiency is achievable in double auction markets even if agents act randomly within their budget constraints. They performed a series of experiments with humans and computational agents who take decisions on a random basis. They referred to these agents as *zero intelligence* agents. In their experiments, they obtained a remarkable collective efficiency with these agents in that by simply applying a budget constraint to the zero intelligent agents, the efficiency in such a market is almost equal to the efficiency in markets with profit motivated humans.

Their results suggest that the achievement of high levels of collective efficiency under classical conditions may place minimal demands on individual rationality, no maximization and not even bounded rationality is necessary. Perhaps the main issue then is not how much rationality there is at the micro level, but how little rationality is sufficient to generate macro-level patterns in which most agents are behaving as if they were rational (Axtell, 2003). We seek an alternative methodology that leaves room for the improvement of the collective system through learning as a substitute for *individual rationality*. Adaptation and evolution may affect the dynamics of a collective system and lead it to evolve to a more efficient outcome.

<Individual learning and social learning> One important issue is the level at which learning is modeled. The two basic possibilities are the individual level and the collective level. Various studies to clarify an essential difference between *individual learning* and *social learning* have been performed. Vriend (2000) and Arifavoric (2004) make these two learning processes more precise. They consider a population of agents who produce homogeneous goods in an oligopoly market. Each firm learns the proper production level the generic algorithm. The first is as a model of social (or population) learning. Each individual agent in the population is characterized by an output rule, which is a binary string of fixed length, specifying simply the agent's production level. The measure of success is simply the profits generated by each rule. The underlying idea is that firms look around, and tend to imitate, and re-combine rules of other firms that appeared to be successful. The more successful these rules were, the more likely they are to be selected for this process of imitation and re-combination.

The second way is to use a model of *individual learning*. Instead of being characterized by a single output rule, each individual agent now has a set of rules in mind. In each period, only one of these rules is used to determine its output level to the market. The rules that were more successful recently are more likely to be chosen. In individual learning, instead of examining how well other agents with different rules did in previous periods, each agent checks how well it did in previous periods when it used these rules itself.

They showed that the individual learning model converges close to the Nash equilibrium output level, whereas the social learning model converges to the competitive equilibrium output level, where no firm gains profit. The difference to modeling learning between these two approaches is often neglected, but they claim that for a general class of games this difference is essential.

1.5 Evolutionary Design of Desired Collective Systems

This book deals with an important question. In collective systems where many agents are all adapting to each other and the collective outcome is extremely hard to predict, what actions should agents take? When there are a huge number of agents, and numerous interactions, a great deal of learning is an attempt to imitate the success of other agents, the resulting collective outcomes are hard to predict. There are also curious questions about how complex systems work and how they can be made to work better. However, no natural method has been proposed for systematically studying these issues. We need to identify and redesign the microscopic rules of agents that produce desirable outcomes at the macroscopic level.

The emerging research field of collective systems represents a fundamentally new scientific approach. In dealing with the above problem we de-emphasize traditional scientific goals such as optimization, equilibrium analysis, and control, in favor of appreciating the importance of emergence, self-organization, diversity, adaptation and evolution.

A collective system is characterized as a system that consists with many learning agents who adapt to other learning agents. A collective system consists of individuals who are learning about a process in which other members are also learning. Learning the true state of the system is therefore quite unlike learning the values of parameters that govern a physical process, for example, or even the parameters that describe a social process that is external to the observer (Young, 2005). When the observer is a part of the system, the act of learning changes the point to be learned. It is therefore unclear whether there behavioral rules of any degree of complexity that can solve this problem consistently. It is also unclear whether the problem can be solved using fixed learning models that bear some resemblance to actual learning behavior in human. Therefore, in order to investigate the performance of a collective system of adapting or evolving agents, we need to explore a new method beyond the conventional equilibrium analysis that emphasizes the dynamic and evolving aspects of the system.

The priority for a desirable collective outcome is stability, which is to be crudely modeled using the idea of equilibrium of the system. However, the condition of stability is not enough, and we need other criteria, efficiency and equity. In the field of economics, efficiency means that nothing gets wasted. This follows Pareto-optimality in taking the absence of waste to be equivalent to the requirement that nobody can be made better off without someone else being made worse off (Binmore, 2001). Efficiency represents the measure of the desirability of collective at the macro level. On the other hand, equity stands for the measurement of the desirability at the micro level.

As shown in Figure 1.2, given a collective system, there is an associated inverse design problem, i.e., how to configure or modify the components (agents) of the system so that in their pursuit of their own interest, they also optimize the global performance. Solving this inverse problem may involve determining and modifying the number of agents and how they interact with each other and what degree of freedom each agent has.

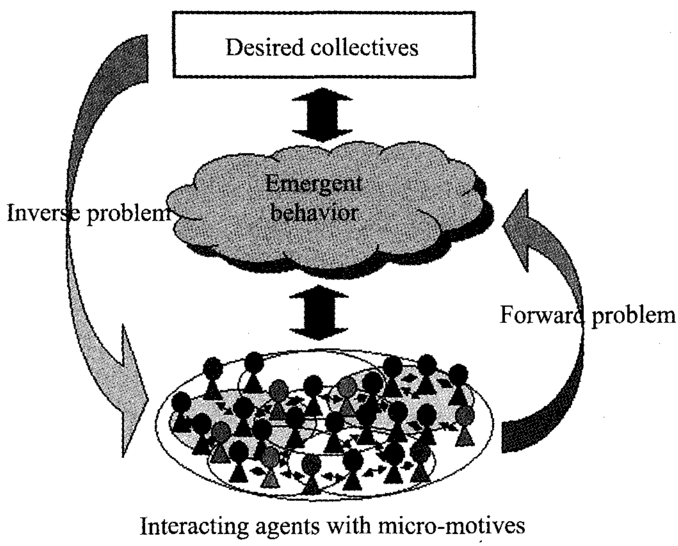

Figure 1.2 The forward and inverse problems of a collective system

In this book, we develop a game-based approach for designing collective systems. For studying frequency-dependent natural selection, game-theoretic arguments are more appropriate than optimization algorithms. This allows the design of agents that have the ability to correct and improve agents' behavioral rules. We also propose a flexible learning paradigm that allows agents to identify situations in which their behavioral rules fail or can be enhanced and to respond by initiating learning processes of successful agents.

Agents interact with each other in complex ways, and linked agent behaviors are highly nonlinear. However, we investigate the situations in which they succeed in organizing themselves into a coherent behavior and produce desirable outcomes through a simulation study. In particular, we focus much of the hidden knowledge of the mysteries of collective behavior of linked learning agents.

Learning agents play the underlying game repeatedly by starting off with a small set of sample rules to be tested. Individual learning then involves, (1) assignment of a rating to each of the rules on the basis of experience, and (2) invention of new rules to replace those rules that end up with a low rating. The rating of a rule is merely the average of the payoffs received when it is used against the opponent. The genetic algorithm uses these ratings as fitness and generates new rules accordingly. Evolutionary learning also follows numerous rules that are causally dependent on previous interactions and on their stored rules. There is no social learning such as imitation or exchange among agents.

However, in the standard model of individual learning, agents are viewed as being genetically coded with a strategy and selection pressure favors agents that are fitter, i.e., whose strategy yields a higher payoff against the average payoff of the population. Natural selection operates on the local probability distribution of strategies within the repertoire of each individual member. An individual's ability to survive and grow is based on advantages that stem from core competencies that represent evolution.

The envisioned research object is quite novel since it requires harmonic and synchronized interactions among self-interested agents. An important aspect of collective evolution is the learning strategy adapted by individuals. In this book, the concept of *collective evolution* can be extended beyond the boundaries of a single agent. Collective evolution is valuable in a social context because it can help to expose new behavioral rules and spread them to other agents who cannot effectively make the proper choice. Each agent is modeled to learn the coupling rule rather than a specific behavior.

Collective evolution also provides a microscopic foundation that is missing in evolutionary approaches. To be sure, we do not see collective evolution as a potential replacement for evolutionary theory, but I argue that they may provide necessary complements, or can be seen as capacities that evolved during evolution to increase the learning efficiency of the individual.

Agents constantly improve their behavioral rules. As such, an important requirement for an efficient behavioral rule is that it should be robust. That is, it may not to be replaced by other rules. Significantly, if a rule achieves both efficiency and equity then it is robust. Our framework synthesizes the principle of collective learning and the mechanism of collective evolution into a coherent approach to the design of the desired collectives. It also provides a device for channeling the complexity of collective systems into manageable and desirable change. Collective evolution is also a driving force for building networks of sustainable interaction that foster stability, efficiency, and equity in collectives of selfish agents.

1.6 Outline of the Book

The foundations of this book lie in three distinct fields: computer science, game theory, and complex theory. These three traditions have made important contributions to this discussion and to the mechanisms proposed to solve the problem of collective action. These traditions will be the subject of this book.

From computer science come insights about collective systems with many agents that can be designed to work together and adapt to each other. Two areas of computer science are important. First, there is the rapid growth of multi-agent systems, which has led computer science into deeper analyses of what it takes for systems of many agents to work together by establishing efficient coordination. Second, there is the field of evolutionary computation, which has fostered an engineering approach to evolution and adaptation. With an engineering approach, one asks how systems can be designed to become more effective over time. By making evolution an engineering problem, evolutionary computation has shed light on how collective system can evolve to a desirable outcome.

Game theory provides insights into how agents can choose actions to maximize their utilities in the presence of other agents who are doing the same thing. A primary question in the recent study of game theory is how each individual should learn in the context of many learners.

Complex systems deal with systems composed of many interacting particles. Complex systems often result in features, self-organization and emergent properties, which are properties of the system that the separate parts do not have. Therefore, the emerging collective outcome is extremely hard to predict. The generation of collective systems with advantageous behaviors beyond our manual design capability requires long-term incremental evolution with continuing emergence.

This book provides some fundamental and common problems for studying adaptation and evolution in collective systems and highlights the benefits and shortcomings of the many related fields. This book also provides some of the essential questions that need to be addressed if the new research field of collective systems is to mature into a new *collective science*.

In Chapter 2, we review the basic concepts of game theory and evolutionary games. Game theory is devoted to the logic of rational decision-makings in social contexts. It is about what happens when self-interested agents interact. The outcome is explained by the concept of equilibrium. Evolutionary game theory, instead, assumes that the game is repeated many times and asks which strategies can survive in the long run.

In Chapter 3, we formulate social games in which there are a large number of agents, each of which faces a binary decision problem with externalities. The outcome depends on the strategy choices of all agents. Fortunately, in certain strategic situations, interactions among many agents can be analyzed by decomposition into the underlying 2x2 games.

In Chapter 4, we consider a global adaptation model of heterogeneous agents. We obtain the relationship between agents' heterogeneous micro-motives and the macroscopic behavior. In particular, we characterize the gains from heterogeneity of agents.

In Chapter 5, we characterize the gains from heterogeneous interactions by formulating knowledge trading in a population. The main concern is in what circumstances knowledge trading can be accelerated by self-interested agents.

Chapter 6 presents a comparative study of two adaptive populations, one in a global environment, and the other in a spatial environment. We also show that the gain from heterogeneity depends on the type of interaction and the location of heterogeneous agents.

In Chapter 7, we study a model in which agents can select partners with whom to interact. An agent needs to select her neighbors to interact with and faces a tradeoff between joining a neighborhood where most agents share her preference or another neighborhood where they have different preferences than hers. Unlike some approaches in these fields, we are primary interested in the problem in which the preferences and even the identities of the agents can evolve over time, rather than situations in which the agents and their preferences are fixed.

In Chapter 8, we deal with social congestion problems. Dispersion games provide a simple model for understanding the mechanisms behind many paradigms for all types of congestion that may arise when we need to utilize limited resources. We introduce a new adaptive model based on the give-and-take strategy, in which agents yield to others if they gain, and otherwise randomize their actions.

In Chapter 9, we explore an alternative learning model, coupled learning, and focus on coupling dynamics that may change in time according to coupled behavioral rules. We show that collective learning of coupling behavioral rules serves to secure desired outcomes by establishing sustainable relationships.

In Chapter 10, we consider another type of social interaction in which agents should be dispersed. In particular, we focus on the emergence of synchronized behavioral rules that sustain efficient and equitable outcomes. Collective evolution reconciles individual-based evolution and leads to socially desirable outcomes.