Introduction

One of the hardest things to understand about humanity is how it is that we can be such sociable creatures, and at the same time, so prone to destructive and anti-social forms of behaviour. Our sociability extends far beyond the obvious fact that we rely upon extremely complex forms of cooperation in order to secure our physical survival. Most of us spend our entire lives embedded in a dense web of social relations, which we depend upon both psychologically and emotionally. When we get lonely, we go looking for company and conversation, deriving pleasure from the simple act of talking with another person. But at the same time that we are so powerfully bonded to one another, we also routinely engage in activities that benefit us at the direct expense of others. The problem is not just with overtly aggressive behaviour. People can be extremely uncooperative in very routine affairs, often refusing to set aside their interests even when the failure to do so makes everyone worse off in the end.

Immanuel Kant suggested that understanding the nature of this latent antagonism was the key to understanding the development of human society:

By antagonism, I mean in this context the unsocial sociability of men, that is, their tendency to come together in society, coupled, however, with a continual resistance which constantly threatens to break this society up. This propensity is obviously rooted in human nature. Man has an inclination to live in society, since he feels in this state more like a man, that is, he feels able to develop his natural capacities. But he also has a great tendency to live as an individual, to isolate himself, since he also encounters in himself the unsocial characteristics of wanting to direct everything in accordance with his own ideas. He therefore expects resistance all around, just as he knows of himself that he is in turn inclined to offer resistance to others (Kant 1784, 44).

This tension between our social and antisocial tendencies has, for subsequent generations of philosophers and social theorists, given rise to a number of extremely difficult technical questions. For centuries, they have been puzzled, not just by how social order in human societies is achieved, but how it could even be possible. In some ways, we know more about how the
activities of ant colonies are organized and reproduced than we do about how human societies function. Human beings, along with colonial invertebrates and social insects, are what evolutionary biologists refer to as an *ultrasocial* species (Wilson, 1974). We sustain and reproduce life through unusually extensive and complex systems of cooperation. In fact, human beings have recently begun to overtake ants, termites and bees as the most social species on the planet (in terms of the number of individuals and the degree of complexity of our societies [Richerson and Boyd, 1998]). This kind of cooperation is uncommon. Far more typical of natural patterns is the level of sociality exhibited by our closest primate relatives, who tend to congregate in tribes of no more than one hundred individuals, and who engage in very limited forms of altruistic behavior.

Apart from being uncommon, there is also something mysterious about the specific form that ultrasociality takes on in human societies. In the case of both colonial invertebrates and social insects, there is a very clear biological and genetic basis for the peculiarly high levels of sociality that they exhibit. With social insects (ants, termites, bees, wasps), for example, it is a legacy of the reproductive pattern known as “haplodiploidy,” which increases the level of relatedness among certain members of the hive or colony. There seems to be nothing comparable in the reproductive biology of humans that could serve as an explanation for our ultrasociality. With respect to reproduction, and in many other aspects, we are not all that different from apes. Furthermore, our dependence upon cultural transmission of these cooperative systems shows that we are not “hardwired,” or programmed to interact in these ways. Given this instinctual underdetermination of behavior, it is unclear how these stable patterns of social interaction are possible at all. According to sociologists Peter Berger and Thomas Luckmann:

The human organism lacks the necessary biological means to provide stability for human conduct. Human existence, if it were thrown back on its organismic resources by themselves, would be existence in some sort of chaos. Such chaos is, however, empirically unavailable, even though one may theoretically conceive of it. Empirically, human existence takes place in a context of order, direction, stability. The question then arises: From what does the empirically existing stability of human order derive? (Berger & Luckmann 1966, 51).
But at the same time that human societies exhibit a greater level of order than our biology might lead one to predict, we are also capable of doing things that, while not producing total chaos, do involve catastrophic failures of cooperation or coordination. Often these failures are not induced by external events, but arise entirely as a consequence of actions taken by individuals within the society. Civil war is the most conspicuous example. This is something that also sets us apart from the ants. In human societies, elements of the social structure sometimes fall apart, and often no one is quite sure why (in part, this is because no one is sure what makes them hang together in the first place). When we refer knowingly to the “rise and fall” of empires, we are tacitly suggesting that every form of social organization, no matter how successful, eventually succumbs to its own inner tensions. There appear to be both centrifugal and centripetal forces at work in every society, but the nature of these forces is very poorly understood.

One of the theorists who has contributed the most to our understanding of these issues is Karl Marx. While previous thinkers had tended to focus on the way that society creates a harmony of interest among individuals, Marx’s focus on conflict and revolution led him to look more closely at the kind of issues that divide them. What he arrived at was the seemingly paradoxical conclusion that the benefits society provides to its members are precisely the source of the conflicts among them. What Marx saw was that the basic institutional structure of society has the function of securing cooperation among individuals (a view captured by the slogan “all labour is social labour”). However, this cooperation, insofar as it is beneficial, generates a surplus – something more than the sum of what individuals could produce on their own. But precisely because this surplus is generated through cooperation, there is no “natural” way of dividing it up among the individuals who helped produce it. This leads to conflict, as individuals compete with one another for a share of this cooperative surplus.

The profound insight here is that cooperation is what produces the conflict. To see this, consider a hypothetical state in which isolated individuals each hunt for their own food. Everyone has certain natural abilities, which, along with a certain amount of luck, determines how much food each will have to eat. There is in this state a natural distribution of wealth – every individual gets what he or she can catch. This means that, in an important sense, there is nothing to fight about. But if these individuals discover that they can hunt more effectively by
going out in large parties, they are then in a situation in which they can produce a cooperative surplus. Hunting together, they can catch more than the sum of what each could catch alone. However, it is now extremely unclear how much of the joint product each individual is entitled to. At the end of the day, instead of having separate piles of food sitting in front of each individual, there is a very large pile of food sitting in front of the group. Even though agents had a common interest in cooperating to increase the size of the surplus, they are now faced with a conflict of interest as they sit down to divide it up.

The cooperative venture gives rise to a common interest because, in principle, everyone can benefit from the arrangement. But the distribution problem generates a conflict of interest because, in principle, any gain for one person is a loss to some other. This can be represented graphically as in Figure I.1, for a simplified society containing only two individuals. Any point in this space represents a possible social state. Each point is an ordered pair, e.g. \((x,y)\), indicating how many units of food player 1 and player 2 get, respectively. Changes in the social state can be represented as movement from one point to another.

We can imagine any particular social state intersected by two axes, one southwest-northeast (SW-NE), the other southeast-northwest (SE-NW). Any movement of the social state along the SE-NW axis creates a pure conflict of interest. Any food that one player gains through this movement is food that the other loses (a “win-lose” outcome). Because the location of the point on this axis determines how much of the cooperative surplus each player gets, this will be called the distribution axis. On the other hand, any movement on the (SW-NE) axis creates a pure commonality of interest. Movement to the NE represents an increase in the total amount of food that each player gets (a “win-win” outcome), while movement to the SW decreases the amount that each gets (a “lose-lose” outcome). Because the location of the point on this axis determines the size of the cooperative surplus, it will be referred to as the production axis. Movement to any point not on one of these axes represents a “mixed” outcome, which can be thought of as the combined outcome of a particular production and distribution decision. Under any mixed outcome, there may be winners, there may be losers, but there will be no symmetry between the size of the wins and loses. Some may gain a lot, while others gain only a little; some may lose a little, while others gain a little, and so on.
Selecting a social state can be extremely difficult. Any point that is selected necessarily involves some kind of a choice about how the distribution problem is to be resolved, and distribution always involves a conflict of interest between the parties (even if they have cooperated in order to produce the surplus that is to be distributed). Finding principles that can resolve these problems can quickly become complicated. Consider the example of the hunting party again. Someone might reasonably suggest that everyone get exactly the same amount. However, another might argue that those with greater natural talent are ill-served by this arrangement, and so everyone should get a share that is proportional to the relative magnitude of the amount that she could have caught on her own. Someone else might then suggest that each individual get what she could have caught on her own, plus an equal share of the surplus. These suggestions might go on and on. Some kind of principle is needed to solve the problem, but it is not clear where we should be looking to find it. Since there is no “natural” way to divide things up, we are, in a sense, on our own. We create the surplus, and so we have to decide what to do with it.

Marx thought that this distribution problem, embedded in every social institution and every system of cooperation, was responsible for all major social conflict. He argued that the basic institutional structure (or mode of production) of every society serves a dual function. Not only does it secure the cooperation needed to generate aggregate gains, it will also contain some mechanism to distribute these gains out to individuals. This is the central idea of what is called
the distributive view of social institutions. In principle, this view applies not only to obviously “economic” arrangements, but to any institution that secures social cooperation – from “big” institutions like the family, the education system and the state, to “little” institutions like the ones that determine how we line up at the grocery store, what time we go to work in the morning, how exams are graded, and so on.

Marx thought that the kind of institutions a society has, and as a result, the way that these institutions distribute out the cooperative surplus, will be determined by a struggle for power among members of the society, rather than by any sort of principled resolution of the conflict. Since systems of cooperation benefit everyone, even when the benefits are distributed out unequally, it will typically be in people’s interests to settle for some type of stable arrangement, rather than see all the potential gains from cooperation destroyed entirely through fighting. Thus after a revolutionary period, in which characteristics of the dominant institutions are “up for grabs,” people will eventually become tired of constant uncertainty and become willing to settle for the best deal they can get. Thus society will settle into a particular mode of production, which, because it is the product of exhaustion, rather than deliberation, will typically reflect the background power relations within society. Social groups or “classes” who gain advantage in such a struggle can entrench themselves by arranging social institutions in a way that systematically favours their interests over those of other groups.

Once these arrangements are settled upon, most people will find that it is no longer in their interest to rock the boat. Even if they are getting a bad deal from society, a bad deal is usually better than no deal. Thus a mode of production that involves significant exploitation can be perfectly stable, in Marx’s view. The conflict over distribution is still there, in the sense that it has not been resolved. It has simply become latent, because the vast majority of the population no one longer has any real incentive to agitate for a fundamental restructuring of society.

The one thing that has the potential to disrupt things, however, is technology. Ultimately, the power that individuals have stems from their capacity to contribute to production (since this determines, among other things, how much the social product will be reduced, should the individual withdraw from the system of cooperation). The technology of production, on the other hand, changes over time, as individual innovations are picked up and reproduced elsewhere. Because this process is cumulative, over time it will change the power relationship between
individuals. (The transition from nomadic hunting to settled agriculture, for instance, dramatically changes the power relationship between men and women.) If technological development is sufficiently dramatic, it may create a mismatch between the institutional structure of society and the underlying power configuration, which will eventually make it advantageous for one social class to challenge the dominant mode of production. When this happens, society enters into a revolutionary phase, all social arrangements become once again “up for grabs,” and a new set of institutions will emerge, one that more closely reflects the transformed relations of power within the society. (In a predominantly agricultural society, those who control the land have the power, thus feudalism emerges as a mode of production reflecting their interests. In an industrial society, those who control the factories and machines have the power, thus capitalism emerges as a mode of production reflecting their interests. It is no accident that it was capitalists, not peasants, who spearheaded the revolution against the feudal aristocracy.)

It is because of this emphasis on technological change that Marx’s view is referred to as historical materialism (Cohen, 1978). However, while technology is the driving force of social change, and provides the directionality of human history, it is the latent conflict over distribution, in Marx’s view, that provides the incentive for individuals to transform social institutions over time. (This is why social structures among human beings change, in a way that they don’t with ants and bees.) Institutions are constantly changing because there is always something for people to fight about, because every form of cooperation automatically generates conflict over how the benefits that flow from this cooperation are to be divided up. This means that in order to understand the history of social change, one must look at the history of struggle over distributive gains. This is why Marx claimed, in the famous opening lines of the Communist Manifesto, that “the history of all hitherto existing society is the history of class conflict” (Marx 1872).

Marx’s vision of a communist society was informed by this view of history. He thought that if it was possible to resolve this underlying distributive conflict, history would come to an end. Since distributive conflict is what drives revolution and social change, solving the conflict would also bring social change to a halt. If we refer to a society in which the underlying distributive conflict has been solved as a just society, then Marx’s view of history amounts to the claim that a just society represents the natural endpoint of social evolution. However, because Marx didn’t believe that a principled resolution to the distributive conflict was possible, he
thought that a just society would only be achieved when technological change had increased the size of the cooperative surplus to the point when distributive conflict would be *trivialized*. Looking at the extraordinary amplification of productive capacity that was occurring in 19th century England at the time that he was writing, Marx anticipated that our ability to satisfy material needs would soon outstrip our capacity to develop new ones. He thought, in other words, that since we (as a society) would soon be enjoying wealth beyond the dreams of avarice, the groundwork was being laid for the emergence of a classless society. It’s not that there would no longer be anything to fight over, it’s just that these conflicts would lose their life-and-death character, and become something more like sport – important only to those who happen to care about such things.

It is now generally accepted that Marx and his followers made a rather catastrophic miscalculation, primarily by underestimating the power of avarice. However, Marx’s basic analysis of social institutions has had a powerful impact on social theorists around the world, and is now so widely accepted that it is often no longer regarded as a specifically “Marxist” theory. Consider, for instance, the following passage from John Rawls’s, *A Theory of Justice*, the most important work in American political philosophy of the 20th century:

Let us assume, to fix ideas, that a society is a more or less self-sufficient association of persons who in their relations to one another recognize certain rules of conduct as binding and who for the most part act in accordance with them. Suppose further that these rules specify a system of cooperation designed to advance the good of those taking part in it. Then, although a society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as by an identity of interests. There is an identity of interests since social cooperation makes possible a better life for all than any would have if each were to live solely by his own efforts. There is a conflict of interests since persons are not indifferent as to how the greater benefits produced by their collaboration are distributed, for in order to pursue their ends they each prefer a larger to a lesser share. A set of principles is required for choosing among the various social arrangements which determine this division of advantages and for underwriting an agreement on the proper distributive shares. These principles are the principles of social justice: they provide a
way of assigning rights and duties in the basic institutions of society and they define the appropriate distribution of the benefits and burdens of social cooperation (Rawls 1971, 3-4).

Marx thought that economic development would effectively eliminate scarcity, and so remove any serious conflict over distribution. Liberal theorists like Rawls, on the other hand, have tried to find a set of principles that will resolve the problem. Once these principles are identified, they can be used to establish the basic constitutional and legal structure of society, and so can provide a satisfactory resolution to the basic source of social conflict. We can refer to a theory of justice as a principle, or set of principles, that allows us to solve this basic problem of social organization, by privileging certain types of cooperative arrangements. A theory of this type could solve the problem in different ways, e.g. it might rank social states from best to worst, or it might specify which social changes are acceptable and unacceptable. A theory of this type functions as a standard that can be used to judge social states. It is not a description of how societies actually function, but rather of how they should function. Thus it is referred to as a normative theory. If you ask the question, “Which social state is best?” or “What outcome is acceptable?” you need a normative standard to answer that question. Of course, such a standard is not entirely “pie in the sky” or wishful thinking, since social institutions are already structured around a set of more-or-less explicit views about the proper organization of society. Thus the major task of a philosophical investigation into the subject is to render explicit, or to state more perspicuously, the normative standards that are already implicit in our best social practices.

In Rawls’s view, Marx was unduly pessimistic about the possibility of arriving at a set of principles that everyone can accept. After all, the problem is not that hard, and people are not entirely unreasonable. The analysis of institutions presented here – and adopted in large measure from Rawls – suggests that a theory of justice should contain two primary normative standards: one that tells us where it is best to be on the common interest axis, and another than tells us where it is best to be on the conflict axis. Some specification of how “tradeoffs” between these standards are to be made will also be required. The normative standard that tell us where social institutions should put people on the common interest axis is normally referred to as efficiency. The standard that tells us where institutions should put people on the conflict of interest axis is
referred to as **fairness**. This gives us two primary ways of judging social institutions – in terms of how efficient they are, and how fair they are.

It is worth emphasizing that analyzing social institutions in terms of efficiency and fairness involves making what social scientists sometimes call “value judgments.” Both standards are designed for judging social institutions from a normative, or moral, point of view. However, it is clear from Figure I.1 that the efficiency standard, because it regulates aspects of the problem in which there is an identity of interests among players, is likely to be relatively uncontroversial. Fairness, on the other hand, because it regulates conflicts of interest, will always be highly controversial. Any outcome selected on the grounds of fairness will be one that each player will have an interest in seeing defeated, because there will always be an outcome that is more advantageous to him or her personally. Because of this difference, there has been a strong tendency among social scientists to regard questions of efficiency as “objective” or scientifically decidable, and questions of fairness as “subjective” or inherently uncertain.

For example, Richard Lipsey, Douglas Purvis and Peter Steiner argue in their classic introductory economics text that, “because it is possible to talk about **efficient** and **inefficient** allocations, but not about **better** or **worse** distributions of income without introducing normative considerations, much of economics concerns efficiency and neglects effects on the distribution of income” (1988, 478). In fact, there is absolutely no basis for drawing this kind of distinction. Both efficiency and fairness are normative standards, both of which can only be derived as part of a general theory of social justice. The fact that one of these principles is likely to be more controversial than the other does not change its basic character. If economists are entitled to make judgments about efficiency, then they are equally entitled to make judgments about fairness. Furthermore, as it turns out, many of the formal techniques of representation that were developed over the years in the field of welfare economics have proven extraordinarily useful when applied to questions of fairness. Thus there has been a growing trend within the discipline of economics toward analyzing issues of fairness (and a growing convergence between the work done by economists and moral or political philosophers interested in such questions).

It should also be noted that, even if questions of fairness are inherently more controversial than questions of efficiency, there is in fact fairly widespread agreement on what the content of these two standards should be (although there is still serious disagreement over how this content
is to be interpreted and applied). The two most common suggestions for these principles are as follows:

- **Efficiency**: We should *maximize* production. The social state should be shifted as far to the north-east of the “common interest” axis as it can be.
- **Fairness**: We should *equalize* distribution. The social state should be as close to the middle of the conflict axis as it can be.

Note that these two standards are in principle distinct. It is possible to move along the distribution axis without affecting the position on the production axis, and vice versa (in concrete terms, one can redistribute wealth without changing the amount, just as one can increase or decrease the amount without changing the distribution pattern). In practice, however, not all social states may be available. This means that trying to satisfy one particular normative standard may involve making compromises with respect to the other. Creating a fair distribution of wealth may reduce the level of efficiency, just as increasing the level of efficiency may create unfairness. In order to solve this kind of problem, it will be necessary to decide which of the two standards should be give priority over the other, or how they should be traded off. It will therefore be necessary, in constructing a general theory of social justice, to look at how these two principles interact when implemented, and to determine some way to resolve conflicts.

The discussion in this text is organized around these two principles. The first half of the book deals with efficiency, the second half with fairness. Both sections start out with an overview and analysis of the attempts that have been made to offer a precise formulation of these principles. Once the most satisfactory formulation of either principle has been articulated, the remaining discussion then examines the way in which our society has attempted to institutionalize it, and the difficulties that have been encountered along the way.
**Key words**

cooperative surplus
distribution axis
distributive view of social institutions
efficiency
fairness
historical materialism
mode of production
production axis
social state
theory of justice