

## Chapter 3. The Moral Economy

Anyone who has ever lived with roommates understands the Hobbesian state of nature implicitly. Roommates quickly discover that buying groceries, doing the dishes, sweeping the floor, and a thousand other household tasks, are all prisoner's dilemmas waiting to happen. For instance, if the food bill is being shared, it gives everyone an incentive to overconsume (because the majority of the cost of anything you eat is born by the others). People also have an incentive to buy expensive items that the others are unlikely to eat. As a result, everyone's food bill will be higher than it would be if everyone did their own shopping. However, if each person buys her own food, then everyone has an incentive to eat everyone else's when they are not looking. When you run out of milk, it's easier to sneak a bit of your roommate's than to go out and buy some more. This makes people reticent to keep large stores of milk in the fridge. As a result, everyone will tend to underconsume, e.g. by buying the smallest quantities available, and keeping as little food around the house as possible. Not only does buying smaller quantities drive up the food bills once again, but everyone must also suffer the inconvenience of running out of food more often. The equilibrium is suboptimal, because of a persistent free-rider problem.

Things are not much better when it comes to other aspects of household organization. Cleaning is a common sticking point. Once you get a certain number of people living in a house, cleanliness becomes a social good. People may keep their own rooms clean, but the shared areas will tend to decline. If it's just your own mess, then the enjoyment of living in a clean environment may outweigh the pain and suffering that it takes to clean it up. But when the mess comes from three or four people, then the effort it would take to clean it up may not be outweighed by the value of a clean environment. Of course, if everyone "pitched in" to clean up, then everyone would be happier. But there is a free-rider incentive – before cleaning, it's best to wait around a bit to see if someone else will do it. Since the person with the greatest tolerance for filth will be able to wait the longest, he will usually be able to exploit the others. In any case, the dishes will stack up in the sink, the carpet will get dusty, and so on. Things will get cleaned less frequently than *anyone* would like, resulting in a suboptimal outcome for all.

But anyone who has lived with roommates also knows that the "state of nature" is not so hard to get out of. People can counteract their tendency to fall into inefficient or suboptimal interaction patterns in a variety of ways. The most common is to *make rules*. For instance, a very

common solution to the usual problems of cohabitation is to draw up a list of household chores, and then assign responsibility for them to different people. If it is decided that the carpet should be vacuumed every week, then someone can be assigned to do that job. If the dishes should be cleaned every night, then people can take turns doing them in the evening. If there is a problem with food pilfering, then one can make it a rule that no one touches anyone's else's food. Setting rules allows people to decide what outcome they would like to achieve, then simply assign the actions to people that are needed to bring it about. A system of rules that designates actions in this way is referred to as an **institution**. The chore list can be thought of as a very simple type of economic institution.

Of course, everyone also knows that just telling people to do things does not automatically translate into a willingness on their part to do them. There is still an incentive problem that needs to be addressed. Thus it is important, in order to understand how institutions function, to understand how they solve the underlying incentive problem.

### 3.1 Internal and External Control

Hobbes realized that rules were the way out of the state of nature. But he was also acutely aware of the fact that getting people to *agree* to a rule is not the same thing as getting them to *comply* with the rule. The fact that everyone agrees to take turns sweeping the floor does not, in itself, give anyone an incentive to sweep it. The old free-rider incentive structure remains intact, so people can be expected to "forget" that it is their turn, or simply to defect from the agreement when the time comes to perform. Of course, this is not what usually happens. Most people can be relied upon to keep up their end of the bargain in cases like this. The question then is how this compliance is secured. How are rules rendered *motivationally effective*?

In considering this question, the first thing that most people pick up on is the fact that the rules are *enforced*. Among roommates, for instance, "forgetting" to do one's chores is usually subject to negative **sanction** – ranging from a rude reminder to being asked to move out. Hobbes turned this commonplace observation into a general theory about how social order is maintained. "Covenants," he said (somewhat dramatically), "without the sword, are but words." He argued that people will obey the rules only if it is in their interest to do so, and the only way to make it in their interest to do so is to take away the benefits of free-riding. This can be done by instituting a sanctioning system that punishes those who break the rules. The way out of the state of nature is therefore to create an authority who can be called upon to enforce agreements that people have

entered into. Hobbes's genius lay in his recognition of the seemingly paradoxical fact that people would voluntarily submit to having their *own will* bound by such an authority. It would be in their interest to do so, because an authority of this type would allow them to avoid suboptimal strategic equilibria.

However, the way that Hobbes tried to cash out this view in terms of a theory of society was remarkably unsuccessful – so unsuccessful that many subsequent theorists have taken it as evidence that his underlying analysis of social institutions was incorrect. The problem is that he was not able to explain how an authority capable of punishing defectors could arise in the state of nature, i.e. out of purely strategic interaction. The problem can be seen very clearly from the simple observation that punishing people usually involves costs for the person who is doing the punishing. So while you can explain rule-following by assuming that transgressors are punished, this simply pushes the problem back one step. What incentive does anyone have to carry out the punishment? Certainly it can be somewhat uncomfortable having to complain or harass one's roommates when they don't do their chores. This gives rise to a new free-rider problem: everyone may put off punishing the person who isn't playing by the rules, in the hope that someone else will do it.

This happens all the time. Consider the familiar situation where someone butts into line in the grocery store, or brings too many items into the express line. Often everyone else in the line will say nothing, but wish that someone *else* would confront him. As a result, the intruder is able to get away with it, because his actions are not sufficiently harmful to any one individual to make it worth their while to sanction him. Punishing defectors and free-riders is a kind of social service, i.e. it generates a positive externality for all those who are harmed by the defection. But because it usually must be carried out by one person, there is a problem motivating individuals to do the punishing. This means that people can get stuck in a suboptimal strategic equilibrium in which they make rules, but then no one punishes those who break them.

Thus there is a problem explaining how social order could be brought about through external sanctions alone. Many social theorists have taken this as evidence that some kind of internal control system is required in order to secure rule-following. The most common view is that people are socialized in such a way that they acquire a very general disposition to comply with social expectations. Properly socialized individuals are willing to play by the rules because failure to do so generates feelings of guilt, remorse or regret. This occurs because individuals internalize the sanctions associated with failure to comply.

According to this view, both external control and internal restraint are essential components of social order. Children are initially taught to conform to the rules by parents and teachers, who systematically reward compliant and punish non-compliant behaviour. (This punishment, incidentally, most often just involves a refusal to cooperate, e.g. parents say "you don't get it until you ask nicely," and so on.) Children come to anticipate these sanctions, thereby acquiring a disposition to follow the specific rule. But they also engage in a process of *generalization*, through which they acquire a more abstract disposition to secure the approval of the parent, or the teacher (Parsons 1951, 210-8). This process of generalization continues through peer-group identification, and the child acquires an increasingly generalized desire to be a "good boy/girl," or a "good student," and later to "fit in," or "play by the rules." With maturity, this disposition is transformed from one of unquestioning conformity to one that seeks to determine independently where one's obligations lie, in terms of what is best for society. At this point, people want to "do the right thing," or "do their part." What is important about this disposition is that it is extremely formal in structure. The desire to "do the right thing" has no content without a set of rules that specify what the right thing to do is. If you promised to meet someone, then meeting him is the right thing to do. If you promised to leave someone alone, then not meeting him is the right thing to do. All of this is a consequence of the rule that says "keep your promises." But the agent may follow this rule for even more general reasons, such as the desire to be "reliable" or to be a "good person." In the limit case, it becomes simply a general desire to act morally, or to "do the right thing."

Naturally, this moral disposition often conflicts with our other desires, and no matter how well socialized we are, no one can resist temptation all of the time. This means that external sanctions are still required among fully socialized adults in order to secure general compliance with the rules. This can be done by making it a specific obligation to punish those who break the rules. More commonly, however, people experience moral outrage and indignation when they witness a transgression of the rules, and so spontaneously sanction the deviant behaviour. This suggests that the disposition to follow the rules acquired through socialization *includes* a disposition to punish those who fail to conform (or at least to ensure that they are deprived of the benefits of their actions.) As a result, agents are punished for engaging in motivated deviance from the accepted set of social rules, and by internalizing this sanction, come to acquire an increasingly general disposition to conform to the prevailing social expectations.

Because the sanctions are internalized by agents, the internal and external components of the social control system are so tightly interwoven that it is often difficult to say where one starts and the other stops. Agents often act from motives that have a hybrid character. Consider, for instance, the person who repays a creditor because he does not want to lose "face" or status. A loss of status is in a sense an external punishment, but the status system only matters because people think it matters. Nothing very bad happens if you "lose face," other than the fact that you are humiliated by very fact that you have lost face. A status system is external, in the sense that the only thing that matters is what others think of you. So if no one knows what you have done, then you can "get away with it." But a status system is internal in the sense that nothing external happens to you if you break the rules, the only punishment comes from your own sense of humiliation. (The "honour" system is slightly more internal, as people will often refuse to do dishonorable things, even if there is no chance of getting caught. An "ethical code," by contrast, is a fully internalized control system, in the sense that it does not matter what anyone else thinks.)

A rule that is followed through a combination of external control and internal restraint is referred to as a **social norm**. The analysis developed here suggests that, in many situations, Pareto-optimal outcomes can be brought about only by instituting social norms that constrain the set of actions available to participants. In the classic prisoner's dilemma, for instance, the two criminals might be friends, and so subscribe to the rule that "friends don't rat out friends." As a result, they might both refuse to confess, even though it is in each of their interest to do so. The consequence, however, will be that they achieve the Pareto-optimal outcome, and minimize the total jail time that they serve. A more formal model of this deliberative process is presented in the following section.

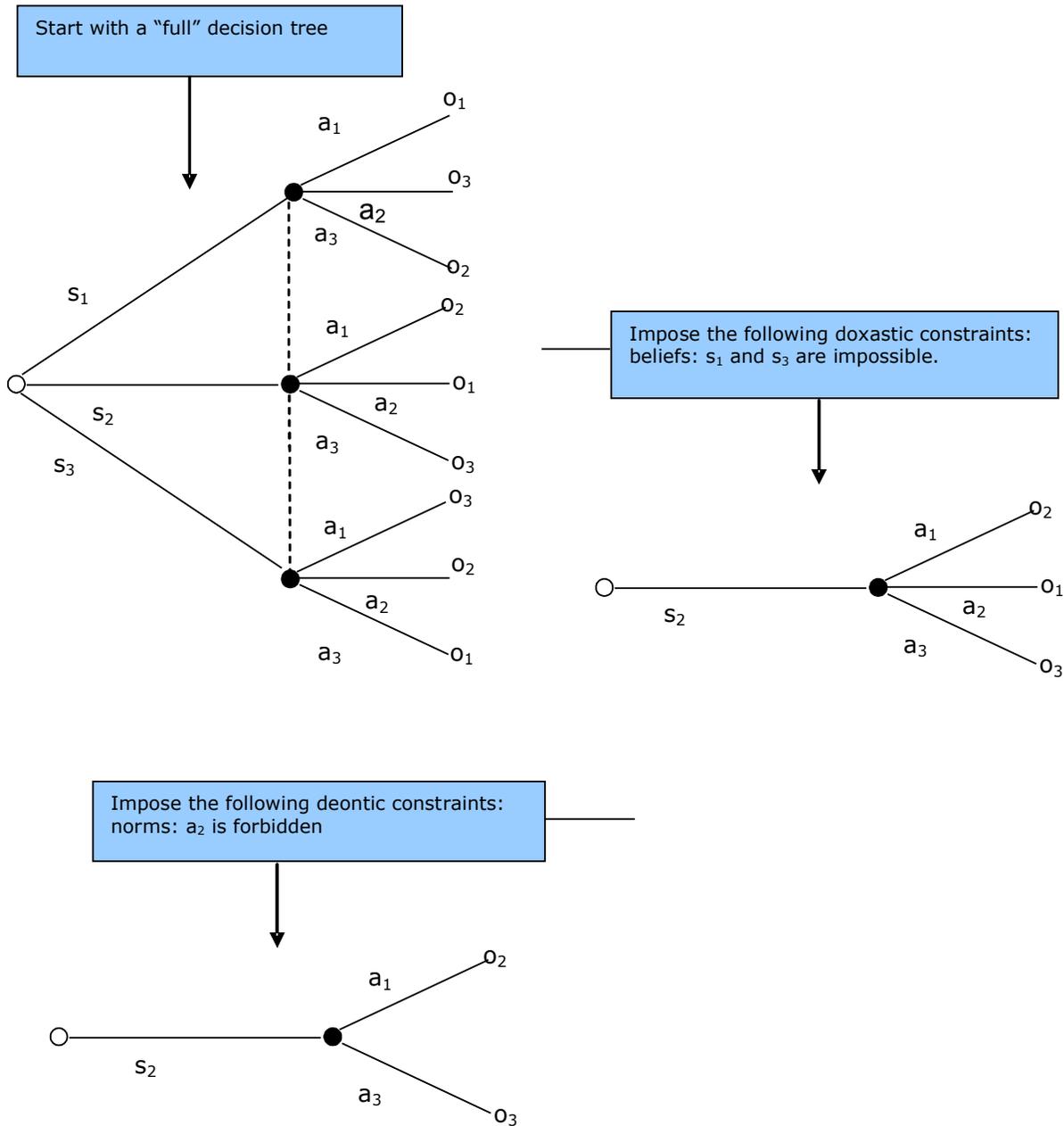
### 3.2 Social Norms

There is a very simple way in which the concept of a moral choice disposition can be integrated into the theory of action developed in the previous chapter. Instrumental deliberation proceeds subject to a set of **doxastic constraints**. This means that we start with the set of possible action-outcome combinations, and then eliminate those that we do not *believe* can occur. Thus our beliefs, in this context, function as constraints on the range of options that we can entertain (hence the term doxastic constraint – *doxa* being the ancient Greek word for belief). We tell agents, in effect, to maximize expected utility, subject to the constraint that their expectations not violate their

beliefs. It is the exercise of these doxastic constraints that generates the *naturalistic pruning* of the decision tree shown in the previous chapter, Figures 2.1 and 2.2.

One way of thinking about the role of social norms in deliberation is to treat them as analogous to beliefs. According to this view, social norms function as **deontic constraints** (from the Greek word for duty – *deon*) that are imposed in a manner quite similar to that of their doxastic cousins. We start out by looking at the options that are *actually* available to us. This amounts to a naturalistic pruning of the decision tree. We then consider which among these options are *morally* permissible. This amounts to a *deontic pruning* of the decision tree, in which we eliminate all the options that violate the prevailing set of social norms. Once this is finished, we can select the action that gives us the most desired outcome.

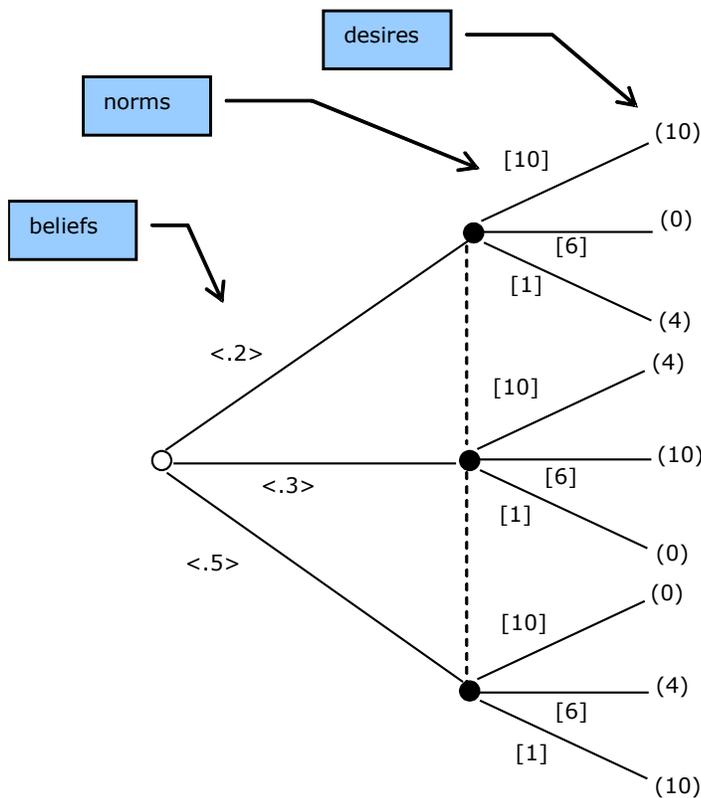
It has long been observed that doxastic and deontic constraints have a similar logic. In the same way that a state of affairs can be either possible, impossible, or necessary, an action can be permissible, forbidden, or obligatory. In practical deliberation the agent begins by determining which of her actions are *possible* ways of attaining a desired outcome. The agent then decides which of her actions represent *permissible* ways of attaining that same outcome. (With respect to the terms introduced above, this is equivalent to cutting out all the actions that are impossible, then removing all those that are forbidden.) She then makes a choice from among the options that remain. The procedure is shown diagrammatically in Figure 3.1. (It does not matter in what sequence the deontic and doxastic constraints are applied. But since it is a waste of time to normatively regulate impossible states of affairs, there is no real point to applying the deontic constraints first.)



**Figure 3.1 Doxastic and deontic pruning**

The style of reasoning illustrated in Figure 3.1 is again somewhat simplified, in that it assumes that states are known with certainty, and that actions are forbidden categorically. But just as states can be known with varying probabilities, actions can be more or less frowned upon. Breaking a promise, for example, is highly undesirable, but the norm can be overridden in some

circumstances. For this reason, permissible actions can be ranked as more or less *appropriate*, in the same way that possible states are ranked as more or less *probable*. A forbidden action, according to this view, can be represented as having an appropriateness of zero. All other actions can be assigned some positive value that specifies their appropriateness, relative to the others. These values can be then be integrated into the agent's utility function, so that they act as a kind of "filter" that reduces the value of normatively inappropriate actions relative to more appropriate ones. These normative considerations can be added to the decision tree in the way shown in Figure 3.2. (Appropriateness factors are shown between square brackets [...], given here on a scale between 0 and 10.)



**Figure 3.2 Decision tree with norms**

We now have a complete decision tree: beliefs associated with states, norms associated with actions, and desires associated with outcomes. The norms can be integrated into the agent's

decision function by simply adding the appropriateness of the relevant action to the expected utility of that action:

$$V(a_1) = 10 + U(a_1) = 10 + [(.2 * 10) + (.3 * 4) + (.5 * 0)] = 13.2$$

$$V(a_2) = 6 * U(a_2) = 6 + [(.2 * 0) + (.3 * 10) + (.5 * 4)] = 11$$

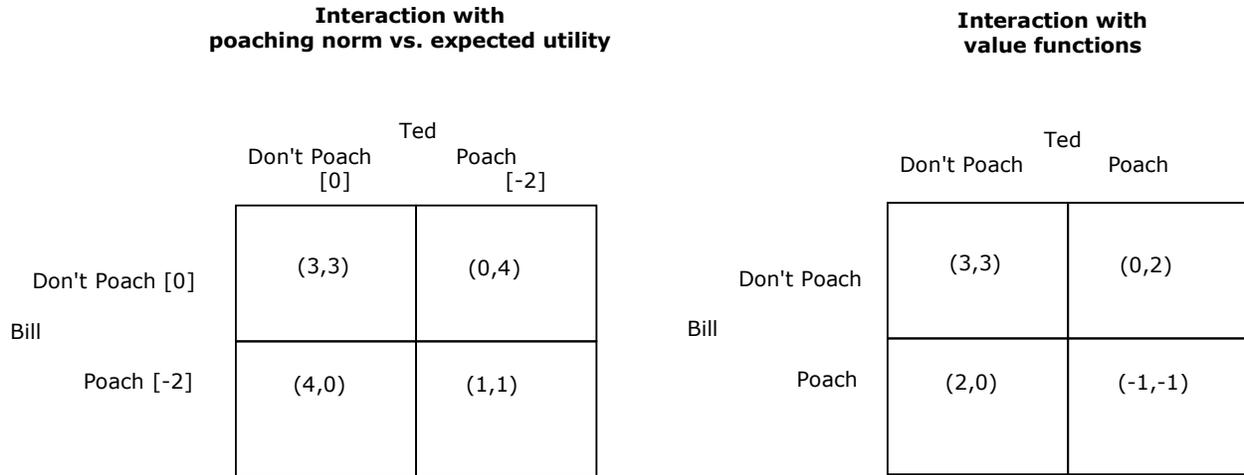
$$V(a_3) = 1 * U(a_3) = 1 + [(.2 * 4) + (.3 * 0) + (.5 * 10)] = 6.8$$

(So  $a_2$ , despite being utility-maximizing, is bumped down through the introduction of deontic constraints.) To avoid confusion, we can call the utility function that results from the introduction of normative constraints the agent's **value function**. It should be noted that the appropriateness of norms and the intensity of desires are both represented here on a scale of 0 to 10. Just as in the case of desires, it does not matter precisely what scale is used to represent the appropriateness of norms. What matters is simply that the numbers reflect the significance of the norms *relative to one another*. It is important, however, that the scale used to represent the appropriateness of norms be correctly calibrated to the scale used to represent desires. If the appropriateness factors range from 0 to 100, and desires only from 0 to 10, then this will mean that the agent considers the most important normative commitment to be ten times more important than her strongest desire. This may be an accurate reflection of how some people weigh these matters, but the precise balance between the two will presumably be specific to the agent in question. As a result, the agent's constrained utility function should make explicit the way the agent weighs normative against instrumental considerations. This can be done by assigning the most appropriate action the same value as the most desired outcome, the least appropriate the same value as the least desired outcome, then multiplying the normative "score" by some value  $k$  that reflects the weight the agent assigns normative considerations relative to instrumental ones. Thus, if  $N(a)$  gives the appropriateness of some action  $a$ , and  $U(a)$  gives its expected utility, the *value* of  $a$  is given by:

$$V(a) = kN(a) + U(a)$$

In many situations, the introduction of a social norm, and hence a deontic constraint, is sufficient to avoid a suboptimal outcome. In the Bill and Ted fishing story, for instance, they could adopt a rule that assigns poaching an appropriateness of -2. Thus the negative value of the action

has the effect of outweighing the expected utility associated with the gains from poaching, as shown in Figure 3.3.



**Figure 3.3. Imposing deontic constraints**

Here all that it takes to eliminate the suboptimal equilibrium is a disposition on the part of the players to assign a reasonable weight to the “no poaching” norm. Regardless of what Ted does, poaching will then have a lower expected value than refraining for Bill, and so he will choose not to. Since the same goes for Ted, both can be expected to refrain from poaching. However, it will not always be the case that a mere decision to respect certain deontic constraints will suffice. In some cases, players must not only respect such constraints, they must also believe that others will do so as well. This occurs in situations where agents are only willing to comply on the condition that others do. In such cases, the deontic constraints of one player generate doxastic constraints for the other. Each must believe that the other will not do what is forbidden. The key difference between this case and the more standard one in which players have beliefs about nature is that the deontic constraints are artificial. While both doxastic and deontic constraints are entirely in the mind of the agents, the agent is *capable* of performing forbidden actions, while he is not able to bring about impossible outcomes. Thus the branches that are deontically pruned remain accessible, or under the control of the agent, in a way that naturalistically pruned ones are not. Furthermore, there will often be an instrumental incentive to take advantage of other players' moral constraint in order to secure a free-rider outcome. This means that deontically constrained interactions have the

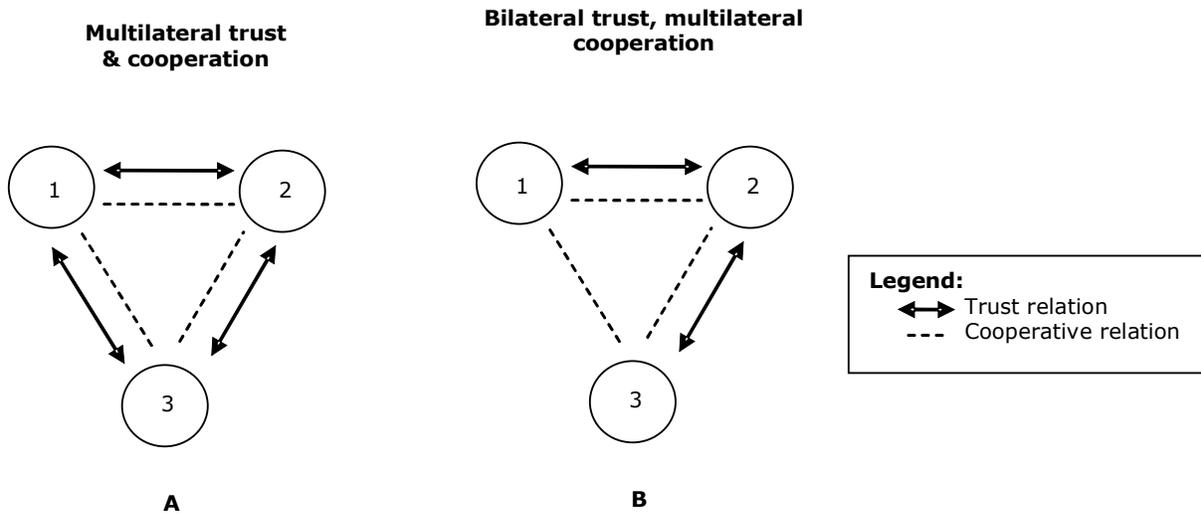
potential to become destabilized, either through outright defection, or the mere anticipation of defection.

These considerations show that the disposition to apply deontic constraints must be matched with *confidence* that others will do so as well. This confidence can be established in either of two ways:

- **Trust:** First, one can attempt to establish trust between the parties. Trust is simply a state in which everyone believes that everyone has a broadly norm-compliant disposition (i.e. assigns a sufficiently high value to  $k$  that their normative commitments outweigh their desires). The best way to achieve this state is if parties are able to somehow reveal to each other that they have such a disposition. This can be done, for example, by establishing a *known history* of having foregone opportunities for profitable defection. People may, for instance, try to determine whether someone can be "trusted to keep a secret," by first revealing a small secret to him, and then slowly building up to larger ones. Trust is best established and maintained in small groups, where interactions are primarily face-to-face, and people can keep track of each other's past history. Often trust is initiated through a default rule that says: trust someone until given reason not to. This is a tendency that is systematically exploited by so-called "confidence artists" or "con men."

- **Sanctions:** In the absence of trust, people can generate confidence that they will comply with the prevailing set of norms by agreeing to be punished in the event that they defect. This is the central insight in Hobbes's view. People want to be able to trust one another, because they can benefit from avoiding collective action problems. This makes it in their interest to submit to some authority who will sanction them in the event of non-compliance. Hobbes's mistake was simply to think that this kind of a sanctioning system rendered trust superfluous. Trust is still required in order to maintain the integrity of the punishment mechanism. However, once this has been established, sanctions can be used to extend the range of cooperative activity beyond the range of individuals who meet and interact on a regular basis. Thus if two individuals do not trust one another, but both trust some third party, then they may arrange with that party to impose the sanctions. In this way, sanctions can be used to extend cooperative behaviour beyond the immediate circle of people who trust one another. Sanctions can in turn help to extend the circle of trust, by giving individuals the

confidence to enter into cooperative arrangements that they might otherwise avoid for fear of being "suckered." Figure 3.4 illustrates the way sanctions can help extend trust relations:



A. Players 1, 2 and 3 all trust one another, all engage in cooperative action

B. Players 1 and 2 trust one another, and players 2 and 3 trust one another, but players 1 and 3 do not. Cooperative action among all three can be sustained if player 2 agrees to sanction 1 or 3 in the event of defection.

**Figure 3.4 Trust, sanctions and cooperation**

These observations allow us to specify more clearly the role that sanctions play in social interaction. People are punished, not because they have performed the wrong action, but because they have violated the *trust* of others. Because relationships of trust are vulnerable to free-riding, but at the same time, permit extremely beneficial forms of social cooperation, trust can easily become a scarce social resource. In particular, people have a strong tendency to defect from a norm as soon as they see evidence that others intend to. The general line of reasoning is: "other people aren't being moral, so why should I?" For instance, a poster in the Toronto subway encourages people not to litter, because *other people* become far more likely to litter once they see litter around them. The littering problem is a standard collective action problem. Most people don't like to ride in a messy subway car. However, given a choice between taking their garbage with them when they get off, or leaving it behind, most people find it easier to leave it behind. After all, since they are

unlikely to return to the same car before it is cleaned, the litter they leave behind is a pure negative externality for the other passengers. In the "state of nature" this means that everyone will leave their trash behind, and everyone will suffer from degradation of the urban environment. The social response to this collective action problem has been to make a rule against littering, which is enforced through both internal constraint and external punishment. However, anti-littering laws are extremely difficult to enforce, and so the results are generally not considered worth the enforcement costs (outside of, say, Singapore). As a result, the anti-littering rule relies very heavily upon voluntary compliance. What the Toronto Transit Commission observed is that people are, by and large, willing to play along with this norm, *as long that they believe that others are too*. However, as soon as they see some litter, they take this as evidence that others are not conforming to the rule. In such cases, they are likely to say "then why should I?," and go on to litter. This tendency is amplified by the fact that many people are actively seeking an excuse to break the rules, simply because compliance carries costs for them in the form of foregone opportunities to free ride. Thus the legitimate desire to avoid being suckered coincides with an illegitimate desire to free ride, giving people an opportunity to rationalize inconsiderate or anti-social behaviour.

The irony is that this kind of "defensive" defection undermines trust just as much as the initial, "offensive" defection. Distrust is a self-fulfilling prophecy. If everyone is disposed to comply with the prevailing norms on the condition that everyone else comply, then there are two stable outcomes: everyone expects everyone else to defect, and so everyone defects; or everyone expects everyone to comply, and so everyone complies. Thus there is a strategic equilibrium and a normative expectations equilibrium. The advantage of the latter is simply that it allows players to achieve a range of outcomes that are Pareto-superior to those obtainable under conditions of distrust. This makes people willing to engage in "experiments" in trust, in order to build up the expectation of mutual compliance.

The general consequence of this view is that morality represents the primary mechanism through which efficient outcomes are achieved in social interaction. This may strike some people as strange. It perhaps sounds less paradoxical to say that in isolated individual action, efficiency is achieved through utility-maximization, while in social interaction, efficiency is achieved through *institutionally constrained* utility-maximization. Examples of this can be seen all around. Queuing is a classic example: people line up for all sorts of things: to get on a bus, to buy tickets, to escape burning buildings, and so on. Everyone in the line has a free-rider incentive to break out of the line

and push to the front. However, when people mob the ticket-counter, or the emergency exit, it tends to slow things down overall, resulting in a longer wait for everyone. The line-up is therefore a perfect example of how moral constraint can bring about a more efficient outcome. The constraint is moral because it is internal, i.e. line-ups are usually self-managing, not explicitly enforced. Their moral character is reflected even in their quasi-ritualistic quality, e.g. people leave a token "placeholder" in the line when they need to step out, and so forth. Furthermore, line-ups are governed by a variety of restrictions, like the fact that you cannot buy a place in line, or sell your place to someone else.

To say, on the basis of this kind of example, that one of the primary functions of morality in society is to secure efficient outcomes sounds like a strange and technocratic misunderstanding of morality. However, this impression fades as soon as one remembers how efficiency is understood according to the Pareto standard. When inefficiencies occur, it means that people are being *harmed*. In an inefficient social state, the well-being of some persons is lower than it could be, even though this reduced level of well-being contributes nothing to the well-being of any other members of the society. If morality is understood as a code that protects people against unnecessary harm, then it is easier to see how it could be internally connected to efficiency.

### 3.3 The Origin of Norms

The claim that social norms represent a form of moral constraint, and that moral constraint promotes efficiency, is subject to two immediate objections. First, there are many examples of social norms that we regard as immoral, and second, there are many social institutions that are very inefficient. In order to deal with these two objections, it is necessary to consider briefly where norms come from, and how they are reproduced.

Turning back to the example of the roommates trying to organize chores, it is fairly obvious where the norms come from. They are established through agreement. Everybody sits down at the kitchen table with a list of chores, and some kind of a division of responsibilities is worked out that is acceptable to everyone. Agreement is essential because the norms only work if everyone expects everyone else to comply. If the roommates are unable to come up with something that is acceptable to all, this will make cooperation impossible, and so they will either have to live with the "state of nature" outcome, or else someone will have to move out and find more like-minded people to live with. (Naturally, if there were only one Pareto-optimal outcome, finding an acceptable arrangement

would be easy. Instead, the task is to pick one point on the Pareto-frontier. As a result, the fairness of each proposed arrangement will be the central issue.)

Things get a bit more interesting, however, once a set of rules has been agreed upon. Social institutions have a tendency to take on a life of their own – to persist despite changes in personnel. To see how this works, consider what happens when one of the roommates, after living under a particular arrangement for a while, decides to move out and is replaced by someone new. The new roommate, anxious to get into the apartment and hoping not to seem difficult, is unlikely to demand a complete renegotiation of the living arrangements. She is more likely to simply accept the established institutions – to take over the old roommate's rights and responsibilities. Even if the rules are slightly unfair to her (e.g. she gets the smallest bedroom, but pays the same rent), the advantages of entry into the cooperative arrangement may easily outweigh any reservations she has over the established distribution pattern. This, combined with the fact that renegotiation is time-consuming and difficult, may make all of the roommates inclined simply to "let sleeping dogs lie."

But through this kind of turnover, a situation can easily develop in which everyone in the apartment adheres to a particular institutional arrangement, even though *none* of them were party to the initial agreement through which it was established. In this way, the reasons for a particular institution can become lost. People may find themselves doing things a certain way, even though they have no idea how they came to be that way. For example, the original roommates may for some reason decide that the vacuuming should be done on Tuesday. Subsequent generations of roommates may uphold that convention, even if they have no reason to do it that way. They reproduce the pattern, however, because each incoming roommate wants to establish *trust* with the others (i.e. not "rock the boat," or come across as a "troublemaker"). The advantages of securing this trust may outweigh the disadvantages that flow from reproducing institutions whose underlying rationale has disappeared (see Kelly, 1998). (To take a real-life example, most Europeans don't know why they eat with the knife in their left hand. There used to be a very good reason. Since their ancestors used to eat with weapons, and since most people are right-handed, picking up the knife with the left hand helped to reassure one's dinner companions that one's intentions were non-hostile. The convention is maintained, even though the reasons for it have disappeared. Now, all that following the convention does is secure trust. If one is dining among strangers, proper use of the utensils helps to reassure everyone that they are in "good company," and so are less likely to be subjected to various forms of inappropriate behaviour.)

If each new roommate simply takes over the role of the one that she replaces, it will eventually create a situation in which an entirely new batch of people are all living together under the rules instituted by the previous batch. Thus the institution becomes independent of the individuals who happen to be sustaining it at any given time. To use an image of Ludwig Wittgenstein's, the institution is like a woven rope which hangs together despite the fact that no single thread runs the full length (1953, 32 [§67]). Naturally, the roommates can get together at any time and renegotiate the arrangements, but there is no reason why this must occur. People will often find it easier just to go along with established patterns, and furthermore, some people may develop a stake in seeing that a particular pattern is reproduced, and so actively resist any pressure to renegotiate.

This kind of persistence over time is a perfectly general feature of social institutions. In society at large, intergenerational overlap is the primary mechanism through which institutions are reproduced. People fall into a certain way of doing things, through piecemeal negotiation and compromise, but this pattern quickly becomes entrenched as new members of the society simply pick up this way of doing things from their parents and elders. Since most of us haven't the time or the inclination to seriously reconsider these practices, most of our institutions are ones that we adhere to simply because we have inherited them in that form. Thus the institutions in any society tend to be independent of the will of the particular individuals who happen to make up the society at any given point in time. Sociologists advert to this phenomenon by referring to the core set of social institutions as the **social structure**, in order to emphasize the way that it remains stable through changes in circumstance and personnel.

This inertial tendency in social institutions is important because it provides the central mechanism through which institutions, even if they start out being efficient and fair, can drift away from that ideal point. The roommates, for instance, may adopt an arrangement in which the person who least minds the vacuuming does the vacuuming, the person who least minds the dishwashing does the dishwashing, and so on. If everyone gets their most preferred chore, then this arrangement will be Pareto-optimal. However, when the dishwasher moves out, all of the remaining roommates will want to keep their original assignments, and so will simply demand that the "new guy" take over the dishwashing. The new guy may accept this request as a cost of entry, even though it is not something that he would have agreed to had he been party to the original negotiations, or that he would agree to if they were to renegotiate the entire arrangement after his entry. Furthermore, if

chores are automatically handed down to incoming roommates, and no one bothers to consult incoming roommates concerning their preferences, then a situation can easily arise in which there are unexploited efficiency gains to be had. The vacuum guy may prefer to wash dishes, and the dishwasher may prefer vacuuming, but neither of them may know it. The institution will still be more efficient than the state of nature, but it will no longer be optimal. Thus social institutions are a mechanism that *permit* agents to achieve Pareto-optimal outcomes, but do not *guarantee* it.

Because institutions have this tendency to go on "autopilot," a variety of devices have been developed that permit the development of more flexible forms of normative control. The most obvious is simply to appoint someone to be the "pilot." Instead of agreeing to a set of rules that prescribe specific obligations, people can instead agree to do whatever some specified individual instructs them to do. This is the kind of agreement that underlies **leadership**. It is ultimately grounded in trust, since individuals must trust the leader to act on their behalf. The leader accepts the responsibility of choosing on behalf of the community, and in return the community agrees to abide by the decisions of the leader. For instance, suppose that the roommates are subject to extremely unpredictable scholastic or workplace time commitments, and so are unable to commit themselves to a rigid timetable for household chores. They might choose therefore to appoint one person as the household chore coordinator, whose responsibility it would be to work out an arrangement of tasks each week that suited everyone's schedule. The coordinator's imperatives would then have the same force as the fixed chore schedule did in the rigid institutional arrangement. Leadership of this type permits much more flexible and responsive forms of normative integration.

When the leader's decision-making power is supplemented with the ability to sanction those who fail to comply, it gives rise to **authority**. Authority structures carry risks along with benefits. While the leader's ability to exercise discretion permits quicker and less costly adaptive institutional response to changes in circumstances, it also creates the potential for abuse. The leader may use this discretion in order to implement arrangements that would not have been agreed to by those affected under more general deliberative conditions. When the leader's decisions are backed by sanctions, they may be implemented even though they are obviously a violation of the leader's fiduciary obligations. However, like social norms, authority structures rest in the end upon the agreement of all affected. While they can be sustained in the face of some deviance and resistance, they cannot run counter to the general interest indefinitely, since the leader must be able to call

upon others to implement his or her decisions, and sanction those who fail to comply. The trust and commitment needed to sustain this loyalty is difficult to sustain in cases where it is clearly not being reciprocated.

Because of the costs associated with renegotiation of institutional arrangements, there is some positive social value in the fact that social actors can resist calls for renegotiation. But social institutions and authority structures can easily get to the point where they become manifestly coercive in character. The frequently-observed fact that strong leaders seem to stay in power long past the point where their net contribution to society is positive may just be a reflection of the fact that every institutional arrangement has a tendency to drift inward from the Pareto frontier as the personnel and circumstances change. It is thus an important component of any social institution that it contain some mechanism to prevent individuals from become too entrenched in their roles.

These considerations show that there is no mystery in the fact that existing social institutions sometimes reflect a considerable departure from our ideals of fairness and efficiency. This does not mean, however, that these institutions are not the primary mechanism -- or even the *only* mechanism -- through which these outcomes can be achieved. It just shows that arrangements which were developed to produce efficiency gains under particular circumstances may tend to become less efficient as circumstances change.

### **3.4 The Moral Economy**

In the beginning, *every economy was a moral economy*. This is one of the primary lessons to be learned from the last century of research in sociology and cultural anthropology. When a group of roommates get together to divide up the household chores, they are unknowingly reproducing one of the oldest and most basic type of economic institution. But while in our society, this form of organization is just a tiny component of a larger economic system that is structured according to very different principles, in pre-industrial societies almost all economic activity is organized through direct normative integration. One of the reasons for this is just that hunter-gatherer and low-tech agricultural economies are not very complex. Under such circumstances, the easiest way to organize production is simply to assign different people to various jobs. Distribution can also be handled by directly specifying who is entitled to what portion of the cooperative surplus. As a result, the "economy" does not form a separable component of the social structure. Everyone lives under a set of rules that specify, with varying degrees of detail, who you can marry,

how to prepare your food, and how you greet your elders, but also, what hunting party you are assigned to, which plot of land you are supposed to farm, how much grain you are entitled to receive, and so on. These rules allow people to avoid a wide range of collective action problems, and so to escape from the "nasty, poor, brutish and short" lives that the unconstrained pursuit of self-interest would guarantee them.

The lack of distinction between "economic" and "social" roles in such societies is reflected in the fact that the science of "economics," founded by Aristotle, is named after the Greek word for household -- *oekonomia*. Aristotle understood economics, or "the art of household management" very broadly, and he included all sorts of practical advice in his economic writings, like how to keep your wife from fighting with your servants, what kind of animals you should hunt, how to make sure your slaves aren't cheating you, and how much wealth you should strive to accumulate. Because the household forms the early locus of productive activity, economic roles often crystallize around kinship relations. The basic division of labour can be established within the household, using production and distribution rules that assign obligations and entitlements to people according to their family position. (A sexual division of labour is usually at the core of such arrangements.) A more complex society can then be constructed by simply adding together more and more of these households.

This kind of arrangement means that instead of having a very complex set of rules that assign individuals from all over the society to specific jobs, there is a simple set of rules that governs action within households. The sociologist Emile Durkheim referred to societies integrated in this way as *segmentally differentiated* (1933, 181) (the idea being that each segment is relatively self-sufficient, and the society is enlarged by adding new segments). The key characteristic of this kind of segmentation is that it provides a mechanism through which a normatively integrated society can become quite large without the normative system becoming overburdened (e.g. imagine living with fifty or one hundred roommates – it might help to break people up into subgroups for the sake of organizing household responsibilities).

Kinship relations supply a very effective set of organizational principles because almost everyone is bound in to the society through some kind of family tie. This means that rules which govern interactions among certain kinds of family relations can serve to organize and integrate the entire society. For instance, among the !Kung hunter-gatherers, everyone in the society contributes in one way or another to the success of hunting parties. However, the game always "belongs" to the

man who made the kill. Distribution is achieved through a social norm requiring that portions of the game be given away to one's relatives:

A man's first obligation at this point... is to give to his wife's parents. He must give to them the best he has in as generous portions as he can, while still fulfilling other primary obligations, which are to his own parents, his spouse, and offspring. He keeps a portion for himself at this time and from it would give to his siblings, to his wife's siblings, if they are present, and to other kin, affines and friends who are there. Everyone who received meat gives again, in another wave of sharing, to his or her parents, parents-in-law, spouse, offspring, siblings and others. The meat may be cooked and the quantities small. Visitors, even though they are not close kin or affines, are given meat by the people whom they are visiting (Marshall 1961, 238; see also Coon, 1971).

This example shows how very complex cooperative outcomes can be achieved through recursive application of quite simple social norms. If everyone in the society is related to everyone else, either by blood or by marriage, then a rule that prescribes a pattern of meat-sharing within families can actually serve to distribute the catch out across the entire society, if it is *reapplied* by everyone who receives a share in the distribution.

One important generalization of this model of distribution is the so-called "gift economy," which is based on the principle of **reciprocity** in exchange. Within such a society, distribution is cut free from specific familial obligations, and is absorbed into the general institution of gift-giving. At the core of this institution is a simple norm stating that whenever someone gives you a gift, you are obliged to repay them at a later date with a gift of comparable value. As a result, gifts can serve as something like an insurance policy in economies that are subject to irregular returns. A hunter, for instance, may make a kill very infrequently, and yet find himself with "surplus" meat when he does. This gives him an incentive to give away as much of the meat as he can to the other hunters, knowing that they will be obliged to repay him when they make their next kill. This distribution system can break the "feast and famine" cycle, guaranteeing everyone a steady supply of meat. In this case, a very simple reciprocity principle allows people to avoid the collective action problems associated with hoarding. By requiring individuals to repay anything that they are given,

it allows them to produce as much of a "social good" as they are able, without having to worry about being exploited.

In both of these examples, the most important feature is that cooperation is secured through moral constraint. According to Karl Polanyi:

As long as social organization runs in its ruts, no individual economic motives need come into play; no shirking of personal effort need be feared; division of labor will automatically be ensured; economic obligations will be duly discharged; and, above all, the material means for an exuberant display of abundance at all public festivities will be provided. In such a community the idea of profit is barred; higgling and haggling is decried; giving freely is acclaimed as a virtue; the supposed propensity to barter, truck, and exchange does not appear. The economic system is, in effect, a mere function of social organization (1944, 49).

People are not forced to distribute out meat to their relatives, or to return a gift, nor is it obviously in their self-interest – narrowly construed – to do so. The entire system presupposes a high level of trust between all participants, and for this reason, requires that everyone in fact be trustworthy. In such a situation, opportunistic or strategic behaviour can only be tolerated within very narrow constraints. For this reason, as Durkheim observed, segmentally differentiated societies tend to have very high levels of social solidarity. The flip side of the coin is that they are able to tolerate very little in the way of dissent or disagreement. Because trust relations are crucial to the reproduction of the society, anyone who challenges the normative consensus poses a danger to the integrity of the community. As a result, the system of norms is usually insulated from critical scrutiny, often resulting in a highly inflexible and traditionalistic society.

It is precisely these high trust requirements that ultimately limit the size and complexity of societies that rely upon direct moral integration in order to secure economic cooperation. For this reason, further population growth requires that the society draw increasingly upon the use of leadership and the application of external sanctions in order to secure cooperation. Thus the emergence of chieftainship, and ultimately the state. In such societies, a central institution is created that overlays the basic reciprocity system with what is called a **redistributive** economy. The leader or chief will demand a portion of the goods produced in the economy (called *tribute*),

and then redistribute it in accordance with some principle. The leader may, for instance, demand a portion of all the rice collected at harvest time, then use it to hold a series of feasts for members of the village, or hold on to it and distribute it back to the farmers for use as seed the following season. The leader might also retain a portion of this tribute and use it to support non-agricultural classes, such as warriors, priests or artisans. Thus the system of central authority permits a specialization of roles that would be very difficult to organize under a system based on reciprocity.

The development of central authority permits the leader to demand tribute not only in the form of goods, but also of labour. Thus the leader may require that each member of the village work a certain number of days out of the year on public projects of his choosing. This makes it possible to provide a wide range of social goods that would not otherwise have been available. For instance, the construction of large-scale irrigation systems in most parts of the world coincided with the emergence of centralized authority. Once these public works acquire a sufficiently central economic role, the state quickly emerges as the hub around which all economic activity revolves. Members of the society receive specific authorization to farm a particular plot of land, or to produce a certain kind of good (from a more-or-less distant member of the leadership class), and in return the state receives a portion of everything they produce. Roles are usually fixed by making social positions among males hereditary, so that sons are obliged to work in the same occupation as their fathers.

Since this kind of society has, historically, been primarily agricultural, the state has often been able to exercise these functions in a very decentralized way. Expansion is generally achieved through the formation of **hierarchies**. Leaders appoint "subleaders" to look after the organizational details in a particular domain. These subleaders may delegate their authority further downward, creating an organizational form with the classic pyramid shape. Again, the more complex form of social organization is generated through recursive application of the basic leader-subordinate rule, and is ultimately grounded in the authority relationship established by this norm.

Classical European feudalism got by with very little in the way of central authority. Chinese feudalism, on the other hand, developed not only an extremely centralized administration, but also pioneered the development of the most highly rationalized form of the hierarchical organization -- **bureaucracy**. A bureaucracy is simply a hierarchy in which members of subordinate classes are able to work their way up to leadership positions. Instead of having a hereditary class monopolize all positions of authority, leaders are chosen from among the pool of subordinates according to

some standard of merit. This gives every member of the society an opportunity to "move up" through the hierarchy, which has very powerful incentive effects. Competition among subordinates to secure advancement reduces the need for external sanctions, which in turn permits the creation of larger and more complex organizations.

Overall, hierarchies require much less in the way of trust among their members than moral economies. They compensate, however, by requiring rather strict *obedience*. The most effective way to ensure this is for the state to claim a monopoly on the use of force. Under these conditions, instead of the system of norms being insulated from critical scrutiny, the state becomes very jealous of its powers and its claim to legitimacy. The state's right to wield power becomes the most sensitive issue – treason replaces blasphemy as the most capital of crimes.

Naturally, in all of these institutional forms, individuals do not disregard their self-interest. The important point is that in a moral economy, people act as deontically *constrained* utility maximizers (in the sense outlined in §3.2) or as value-maximizers, not straightforward utility-maximizers. Individuals will always try to do the best that they can for themselves, but they do so *within* a given institutional structure. No society is able to transform individuals into automata who simply act out social roles. People will always be looking for an "angle" or a way of getting ahead (although this usually takes the form of competition for status, rather than wealth). What makes the moral economy distinctive is that the set of social norms used to constrain the pursuit of individual advantage at the same time serves to *organize* production and distribution. Economic roles are occupied by individuals who are under a specific social obligation to assume those particular roles.

Because the moral economy consists of a set of institutions that generate outcomes that are not strategic equilibria, it remains somewhat precarious. The potential for free-riding is always present. Contrary to what Polanyi says, the fear of shirking does not go away, any more than the fear of theft, disloyalty, or ingratitude. These tendencies are just subjected to very powerful forms of normative control. The danger of "reversion" to an underlying strategic outcome remains latent in any institutional structure, should the normative system lose its motivational efficacy. This is why, throughout human history, people have worried about society falling apart when they see a visible decline in moral standards (or more commonly, when they see a *change* in moral standards, which they interpret to be a decline). This type of concern is one of the most pervasive forms of anxiety in human society, although in one sense it is simply another component of the normative control system. After all, having someone constantly raising the alarm over one or another type of

decline serves a second-order control function – it sanctions people for *failing* to sanction those who violate social norms.

The fact that social structure is an artificial construct, one which lifts human society out of the state of nature through force of will, is reflected in the widespread fear of a lapse into barbarism (think of classic movies like *Planet of the Apes*, *The Road Warrior*, etc.). No one ever seems to worry about barbarians, through some catastrophic shock, "lapsing" into civilization. There is a significant element of *design* in all of our social institutions, and these institutions do not reproduce themselves spontaneously, without some effort on our part. Unfortunately, this fact has been obscured in the last two hundred years by a number of revolutionary innovations in the way that economic activity is organized in our society. At the core of this revolution is the form of social organization known as capitalism. What capitalism – or more precisely, the market economy – permits is the organization of economic cooperation with a kind of *minimal* morality. Markets are an extremely clever innovation. Rather than assigning particular tasks and entitlement to individuals, a market economy assigns them nothing more than a set of property rights. A well-defined system of property rights then allows them to act instrumentally while at the same time avoiding many of the collective action problems that characterize the state of nature. This in turn creates the possibility of a massive decentralization of economic activity and control within the society. As a result, a capitalist society is able to get along with a lot less morality, a lot less normative control, than most societies in history.

### **Key Words**

authority

bureaucracy

deontic constraints

doxastic constraints

hierarchies

institution

leadership

modality

reciprocity

redistribution

sanction

social norm

social structure

value function