Coleman 1990
(from Hechter/Opp/Wippler

2

*The Emergence of Norms**

James S. Coleman

Much of sociological theory takes social norms as given, and proceeds to examine individual behavior or the behavior of social systems when norms exist. Yet to do this without raising at some point the question of why norms exist at all, and how they came into existence, is to forsake the more important sociological problems in order to address the less important.

Whatever the reason for neglect of this question (and the reason differs for different sets of theorists), I will show in what follows that the emergence of norms can be accounted for by two simple principles. The first of these concerns the conditions in which a demand for effective norms will arise; the second concerns the conditions under which the demand will be satisfied. Both sets of conditions may be described as social structural.

As much as any single concept in the social sciences, a "norm" is a property of a social system, not of an actor within it. In part because it bears a correspondence, at the level of a social system, to "value" at the level of an individual, and because norms may affect individuals' values, this concept has come to play an extensive role in theories developed by

*An extension of the work described here can be found in Chapters 10, 11, and 30 of Coleman (1990).

some sociologists. The reasons are even more fundamental. The concept of a norm at a macrosocial level, governing the behavior of individuals at a microsocial level, provides a convenient device for explaining individual behavior, taking the social system as given. It is a device especially useful for those sociologists characterized by Sorokin (1928) as members of the sociologistic school of social theorists. This is a school of which Emile Durkheim was the most prominent member, for he began with social organization and in a part of his work asked the question, "How is an individual's behavior affected by the social system within which he finds himself?"

For another school of social theory, of which Talcott Parsons was the most prominent member, the concept of norm provides a basis for a principle of action that plays a role in the theory comparable to that of maximizing utility for rational choice theory. The principle is approximately, "Persons behave in accordance with social norms," leaving examination of the content of norms as the theoretical task at the macro level.

Quite apart from its role in social theory, the use of the concept of norm in describing how societies function is important. This is especially so for the description of traditional stable societies. A description of the functioning of caste India without the concept of *dharma* (meaning "duty" or "appropriate behavior," or "behavior in accordance with accepted norms"), would hardly be possible.[1] In stable societies, fixed or slowly changing norms constitute an important component of the society's self-governing mechanisms.

Both the evident relevance of norms for the functioning of societies, and the importance of norm as a concept throughout the history of social theory, provide bases for its position in contemporary social theory. It has not one but *two* entries in the *International Encyclopedia of the Social Sciences* (both written by sociologists), one of which begins, "No concept is invoked more often by social scientists in explanations of human behavior than 'norm.'" As an example, Ralf Dahrendorf (by no means one of those sociologists most wedded to the concept) in an essay "On the origin of social inequality," states

> The origin of inequality is thus to be found in the existence in all human societies of norms of behavior to which sanctions are attached . . . (T)he derivation suggested here has the advantage of leading back to presuppositions (the existence of norms and the necessity of sanctions) which at least in the context of social theory may be taken as axiomatic" (1968, p. 104).

Though norms and sanctions may be taken as axiomatic by many sociologists, they constitute an unacceptable *deus ex machina* for other

social scientists—a concept brought in to explain social behavior, yet itself left unexplained. Some rational choice theorists, armed with utility maximization as a principle of action, regard the concept as altogether unnecessary. To take this stance, however, is to ignore important processes in the functioning of social systems, and thus to cripple the theory. It is one thing to refuse to take norms as given, as starting points in social theory; it is quite another to ignore their existence altogether. Here I refuse to take norms as given, but I ask how they can emerge and be maintained among a set of rational individuals.

In the present theory, social norms enter in the following way: They specify what actions are regarded by a set of persons as proper or correct, or what actions are improper or incorrect. They are purposively generated, in that those persons who initiate or help maintain a norm see themselves as benefitting from its being observed, or harmed by its being violated. They are ordinarily enforced by sanctions, which are either rewards for carrying out those actions regarded as correct, or punishments for carrying out actions regarded as incorrect. Those holding the norm claim a right to apply sanctions, and recognize the right of others holding the norm to do so as well. Persons whose actions are subject to norms (who themselves may or may not hold the norm) take into account the norms, with their potential rewards or punishments, not as absolute determinants of their actions, but as elements which enter their decision about what actions are in their interest to carry out.

Furthermore, a norm may be embedded in a social system in a more fundamental way that is internal to the individual, with sanctions applied by the individual to his own actions. In such a case, a norm is said to be internalized, and the individual feels internally applied rewards for actions held to be proper by an internalized norm, or internally applied punishments for actions held to be improper by an internalized norm.

Beyond individual norms there is interdependence among norms, such that many are part of a "structure of norms." The most elaborate such structures are those described by *dharma* in India or analogous systems in other societies with long cultural traditions.

Viewed in this way, the tasks for the theorist are substantial. First, one must establish the conditions under which a norm with a particular content will arise. This includes answering the question of why a norm does not always arise when the existence of an effective norm would be in the interests of all or most persons. Related to this are the tasks of specifying who will come to hold the norm, and specifying whose actions will be its target.

Second, is the determination of the strength of sanctions and their

prevalence, recognizing that applying a sanction may entail costs for the sanctioner. Related to this is the question of what kinds of sanctions will be applied, since there is a variety (and, empirically, it is evident that various kinds of sanctions are in fact applied, ranging from those that damage or enhance reputations to those that impose physical damage or provide material benefits).

Examples of Norms and Sanctions

To gain some sense of what is meant by norms and sanctions, it is useful to begin with examples.

1. A child aged three, walking with its mother on a sidewalk in Berlin, unwraps a small piece of candy and drops the cellophane on the sidewalk. An older woman, passing, scolds the child for dropping the cellophane and admonishes the mother for not disciplining the child. A child aged three and a half, walking with its mother on a sidewalk in New York City, unwraps a piece of cellophane and drops the paper on the sidewalk. An older woman is passing by, but says nothing, not even noticing the action of the child.

Several questions are raised by this example. Why does the older woman in Berlin assume the right to scold the child and admonish the mother? Why does a woman in a similar circumstance in New York not do the same? Does the woman in New York not feel the right to discipline the child, or does her failure to act arise from other sources?

2. In an organization that provides free coffee and tea to its employees, one employee who drinks tea takes his cup to obtain hot water to get some tea. All tea bags are gone. He expresses no dismay, remarking to another person standing there: "This often happens, but I have taken some tea bags back to my office just for such occasions." The other person responds in a disapproving way, "It's people like you, stashing tea bags away, who create the problem."

This example also raises questions. Again, how did the second person come to acquire a right to express disapproval? And why did the first person leave himself open for such a comment, by his remarks? Further, why did he accept the disapproval of the second person, apparently acknowledging the right of the second person to impose this sanction?

3. A high school girl on a date at a beach house on the lake finds herself in a crowd in which the others, including her date, are smoking marijuana. The others encourage her to do so as well, showing disapproval and disdain at her reluctance. The reluctance, in turn, is produced by her knowledge that her parents would disapprove.

This example raises questions about conflict: Can there be two conflicting norms governing the same actions? If so, then what determines which will govern the action, if either does? And if conflicting norms do occur, what is the class of situations in which they arise?

Classes of and Distinctions among Norms

The diversity among the examples above suggests that some definitions and some classifications among norms will be useful at this point.

First, norms are directed at certain actions, which I will call *focal actions*. These are the actions over which persons other than the actor assume rights of partial control. In the first example of the 3-year-old and the cellophane candy wrapper in Berlin, the focal action was dropping the wrapper on the sidewalk, or more generally any action that had the effect of littering the sidewalk.

Some norms such as this one discourage or proscribe the focal action; I will call these *proscriptive* norms. Other norms, such as the norm of smoking marijuana held by the young people at the lake, presented in example 3, encourage or prescribe the focal action (*prescriptive* norms). Norms of the first type provide a negative feedback in the system, damping out the focal action, while norms of the second type provide positive feedback, thus encouraging the focal action. When there are only two possible actions, of course, one is prescribed and the other is proscribed by the same norm. For example, the norm of walking to the right when encountering another pedestrian walking in the opposite direction is simultaneously prescriptive and proscriptive, making this distinction meaningless. The distinction is meaningful only when the number of alternative courses of action is greater than two.

For any norm, there is a certain class of actors whose actions or potential actions are the focal actions for a norm. The statement, "Children are to be seen and not heard," specifies a norm in which children constitute this class. I will call members of this class *targets* of the norm, or target actors. There is also a class of actors who benefit from others' observance of the norm and are potential sanctioners of the target actors. These are actors who assume the right to partially control the focal action, and are seen by others who hold the norm to have this right. In the norm given by this statement, parents or adults more generally are those who benefit. It is possible that children also hold the norm, but its operation and supporting sanctions do not depend on this. I will call

those who benefit from the norm (who are also ordinarily the potential sanctioners) *beneficiaries* of the norm. The current beneficiaries of the norm may be those who initiated it, or they may have merely continued the enforcement of a norm initiated by others who preceded them.

For some norms, like the one for children described above, the targets of the norm are not the beneficiaries. The norm is held by one set of actors and directed toward actions of another set. These will be labeled *disjoint* norms, both because the set of holders and the set of targets are disjoint (or largely so), and because their interests are as well: The beneficiaries have an interest in the norm being observed, and the targets have an interest in the focal action being unmodified by the norm.

For many norms, however, including all those in the earlier examples (except for the norm about their daughter smoking marijuana, held by the parents of the girl at the lake), the set of beneficiaries of the norm coincides with the set of targets. Actors hold a norm about their own actions. Here, the interests favoring observance of the norm and those opposing its observance are contained within the same actors. Each is simultaneously beneficiary and target of the norm—*conjoint* norms.

A clarification of what is meant by the term "sanction" is also useful. If holding a norm is assumption of the right to partially control a focal action and recognition of other norm beneficiaries' similar rights, then a sanction is the exercise of that right. A sanction may be negative, directed at inhibiting a focal action which is proscribed by a norm, or positive, directed at inducing a focal action toward which there is a positive norm. The terms "sanction" and "effective sanction" will be used interchangeably, indicating by either an action on the part of a norm beneficiary that has some effect in moving the focal action in the direction intended by the sanctioner.

One final distinction concerns selection of a focal action from among a set of mutually exclusive actions to be discouraged or encouraged by the norm. In some cases, the focal action is largely arbitrary, while in others it is not. The first is exemplified by the convention of driving on the right side of the road (or in England and Australia, on the left). It is arbitrary whether the action to be defined as "correct" is driving on the right or on the left. Once that convention has been established, however, all are better off if each follows the convention. The interests in a particular direction of action depend on whether that is the one being carried out by others. If it is merely a convention which established the direction of the norm, I will call this a *conventional* norm.[2]

For many norms, the focal action is not arbitrary. The target's interests lie in a direction of action opposing observance of the norm, while the holder's interests lie in the direction of action favoring observance of the norm. These interests in a particular direction of action would remain,

whether or not the norm were in existence, and independent of others' directions of action. In this case, the direction of the norm depends on more than convention. I will call these *essential* norms. This last distinction can be illustrated, as Ullmann-Margalit (1977) has done (and as I will do shortly), by use of simple payoff matrices from the theory of games.[3]

Externalities and the Genesis of Interests in a Norm

Actions that have externalities generate interests in the action among those actors who experience the externalities. Yet there is no general way in which the consequences of the action for the other affected actors can enter the utility function of the actor taking the action. Actors harmed by the action that benefits the actor in control of the action experience negative externalities, as exemplified by nonsmokers sitting near a smoker. Those benefitted by it experience positive externalities, as exemplified by passers-by who benefit from a householder's cleaning snow from his sidewalk. The social problem in the first case is how to limit the action (and how much to limit it) that is harming others. The problem in the second case is how to encourage and increase the action, and to what level. A special case of the latter is the problem of paying the cost of a public good, when each actor's action has beneficial consequences for others, but the benefits to himself are less than the costs he will incur. Only if enough actors can be induced to carry out the action to bring the benefits above the costs for each will the public good be provided. A parallel problem exists for a public bad, as in overgrazing of a commons, in which each herd-owner's expanded grazing will bring him a his own benefit, but at a cost to others. Only if the herd-owners can all be induced to limit their grazing will the levels be reduced to that which will provide maximum nutrition.

When an action generates externalities for others, they may be able to make their interests felt through wholly individualistic means. One of these actors may engage in an exchange with the actor whose action imposes externalities, offering something to bring about the outcome he desires, or threatening this actor with an outcome on another event that goes against his interest. But this may not be possible, if the externalities are spread among several actors, no one of whom can profitably make such an exchange.

This solution is a special case of that introduced in Coase's 1960 paper, "The Problem of Social Cost." The general solution is to develop a market in rights of control, in which the actors who do not have control of the action may purchase rights of control from those who do, limited

only by their interest in the action and their resources. It is easy to see that if there are no transaction costs in such a market, then the outcome will be a social optimum (which is defined only relative to the initial resource endowments of the various parties in the market), at which no further exchanges are mutually beneficial. Those harmed by this level of action would be even more hurt by parting with the resources that the actor controlling it would take to limit it further.

In the case of a public good, each of the actors who is benefitted by the actions of others would exchange rights of partial control of his own action for rights of partial control of the actions of each of the others. For example, they might unanimously vote to adopt a highway speed limit. This in effect would make each actor's action controlled by all the potentially affected parties.

Related markets have been developed in regulation of environmental pollution, not with the amount of total pollution allowed set by market forces, but with marketing of rights to pollute among those who operate pollution-generating plants. And it is likely that the process by which the provision of public goods comes under the control of a collective decision (that is, the shift of activities from individual control to governmental control) can usefully be conceptualized as exchange, among those affected, of rights of partial control over their actions. This results in each having a vote in determining the actions of each.

Yet there are many activities in society in which markets in rights of control cannot easily come into being, for one or another reason, and in which collective decisions are not feasible. In a social situation where one person is smoking and another finds it irritating, the second can hardly come to the first and say, "How much will you take to stop smoking?" Or a high school girl on a date in which all others present would like her to smoke marijuana while her parents would not can hardly ask for bids from the two parties for control of her action. There is a wide range of situations in which an action has extensive external effects, but in which a market in rights of control of the action is either impracticable or illegal.

The first principle referred to at the outset is that interests in a norm arise when an action has similar externalities for a set of others, when markets in rights of control of the action cannot easily be established, and when no single actor can profitably engage in an exchange to gain such rights. Such interests do not themselves constitute a norm, nor ensure that one will come into being. They create a basis, a demand for a norm on the part of those experiencing externalities.

The externalities created by the action may, as indicated earlier, be positive or negative. In high schools, for example, positive externalities are created by athletes who contribute to the success of a team, which in

turn contributes to the school's general standing in the community (which, in turn, contributes to the other students' feeling of well-being or pride). Often a norm does arise, one which encourages potentially good athletes to devote their energies to interscholastic sports.

In contrast, scholars who obtain especially high grades create negative externalities for other students, insofar as the teacher grades on the curve. High-performing students increase the effort necessary to produce the same grade for other students, thus making matters more difficult for them. Often a norm also arises in this case, with students imposing norms to restrict the amount of effort put into schoolwork.[4]

How a norm actually comes into being once a demand is created by externalities is altogether another matter, which will be examined later. But the genesis of a norm lies in externalities of actions that cannot be overcome by simple transactions which would put control of the action in the hands of those experiencing the externalities.

Several points follow from the central premise stated above, that interests in a norm arise when actions have external effects. The implication of this premise is that the potential norm will be held by all those who are affected in the same direction by the action. If a norm does arise, it will be those persons who claim a right to have partial control over the action, and it will be they who will exercise the claim by attempting to impose normative sanctions upon the actor to induce the direction of action that benefits them, though often at the actor's expense. A further implication is that a potential conflict of norms arises when an action has positive externalities for one set of persons, and negative ones for another. Such opposing externalities can be seen in the example of the girl whose friends' approval is contingent on her smoking marijuana, and her parents' is contingent on her not doing so (or their ignorance of her doing so). If she does not smoke, she dampens the party, destroys the consensus, and perhaps reminds some of those present of the normative conflicts they are under. If she does smoke (and her parents learn of it), they are made unhappy as their hopes and aspirations for her are undercut.

The structure of interests created by externalities in which norms have their genesis may be seen more systematically by use of simple situations the outcomes of which can be described by payoff matrices as used in theory of games. Here I will use this device in a context that is examined more fully in Coleman (1990).

Suppose two persons are each told, "You make take either of two actions, contribute $9 to a common project, or contribute nothing. For each $3 that is contributed, an additional $1 will be earned by the project (altogether $4 return for each $3 contributed). Then the total will be divided between the two of you, regardless of who made a contribution."

		A_2: Contribute	A_2: Not contribute
A_1	Contribute	3, 3	-3, 6
	Not contribute	6, -3	0, 0

Diagram 2.1.

Each can assess the net gains or losses for himself and for the other, for each combination of actions of the two. These are expressed in dollars in Diagram 2.1, with the value of the outcome for A_1 and A_2 listed in that order in each cell.[5]

If neither contributes, there is no gain or loss for either. If A_1 contributes and A_2 does not, A_1's \$9 contribution plus the \$3 earned will be divided equally, giving \$6 to each. For A_2, this will be a net gain, as listed in the upper right-hand cell of the diagram. But for A_1, the size of the original \$9 contribution must be subtracted, giving him a net loss of \$3. The same outcome in reverse occurs for the case in which A_2 contributes and A_1 does not.

This situation creates a pair of actions each of which has externalities for the other actor. As Diagram 2.1 indicates, A_1's action of contributing or not makes a difference of \$6 (between 3 and -3 or between 6 and 0) to A_2, and A_2's action makes a difference of \$6 to A_1. Furthermore, in both cases, the externalities move in the direction that opposes the actor's own interests. A_1 is better off if he does not contribute (whichever action A_2 takes), but this makes A_2 worse off. A similar situation exists for the effects of A_2's preferred action on A_1. Finally, the result of this situation is that the external effects of the other's action are greater for each than are the direct effects of his own action. For A_1, his own action only makes a difference of \$3 to him, while A_2's action makes a difference of \$6. Similarly, the reverse is true for A_2.

The result of this condition is that each will have an incentive not to contribute (since one is \$3 worse off by so doing), with the result that each gets \$0. Yet if both contributed, each would gain \$3. The optimal action for each actor gives a social outcome which is not optimum. Both

would be better off if both took actions which were *not* individually optimum, that is, if each contributed to the project.

Much has been written about this structure of outcomes, most of which need not concern us here. (For references to some of this literature, see Axelrod, 1984.) What is of interest is Ullmann-Margalit's discussion of this structure as "calling for" or "generating" one type of norm, which she calls prisoner's-dilemma (or PD) norms. Her argument is that such a structure of outcomes creates an incentive on the part of all parties to have a norm that will constrain the behavior of each toward carrying out the action that is better for the others—in this case to contribute to the joint project. Using the terminology introduced earlier, such a structure of interdependence of actions creates externalities for each and thus an individual interest in the creation of a norm.

However, in situations of this sort, where two persons' actions affect each other in the way shown in Diagram 2.1, a norm is not necessary at all. Either can propose an exchange in which each gives the other rights of control to his action and gets rights of control of the other's action.[6] Each has resources (his own action) that are of more value to the other than the resources held by the other (the other's action). Thus by exchanging rights of control, each gets something that is worth more to him than what he gives up. In exercising that control over the other's action, each does so in the direction which benefits himself, and in so doing brings about a social optimum. A_1 contributes A_2's \$9, A_2 contributes A_1's \$9, and both gain \$3 from the double contribution.

It is always true that with a pair of interdependent actions in which the self-interested action of each imposes negative externalities on the other that are greater than the benefits that the other's own self-interested action brings, a mutually profitable exchange is possible.

Logistics may, of course, preclude exchange. In the game-theoretic analysis of this structure, the possibility of exchange is excluded, because by assumption the players cannot communicate. But no such constraint is necessary here. Because norms can arise only where there is communication, two-person exchange is possible in all those two-actor cases where the possibility for a norm exists. There is an apparent exception in those cases where communication exists before and after the action, but not during the action itself. However, whatever agreements are reached before the action, whatever retributions are taken after the action, need make no reference to a "norm," but can be treated wholly within the framework of bilateral exchange—although possibly of course requiring introduction of notions of trust and mutual trust.

The one true exception, in which the social optimum is neither reached by an individualistic solution nor by a bilateral exchange, is that in which interactions are indeed pairwise, but the two actors are not in

contact both before and after the action (or will meet only in the distant future), and thus have no opportunity either to make an agreement or to carry out the terms of a prior agreement.[7] In that case a norm, in which sanctions are imposed by others who are in contact with the actors after the action, can bring about a social optimum while a pairwise exchange cannot.[8]

Before proceeding, it is wise to clarify what "exchange" ordinarily implies in the current context, for the example may otherwise be misleading. The imagery evoked by exchange in the context of this example is one in which one actor approaches another with an offer, "You let me make your decision, and I will let you make mine," or "Let us contribute together," or something similar. This is certainly what happens in some cases. But in an examination of the emergence of norms, it is appropriate to think of a succession of comparable projects, extending over time, with a new decision taken each time. Then the possibilities expand to bring about exchanges, implicit or explicit, that cover two or more projects (for example, "If you fail to contribute this time, I will not contribute next time"). This possibility is especially relevant for those cases in which it is not logistically possible to exchange control or rights to control on a given occasion. It is also relevant for those cases in which there is not a "project" involving simultaneous contribution, but separate actions of each actor which exhibit the same pattern of internal and external effects. That is, actor A_1 must decide whether to take an action (such as watching his neighbor's house while he is gone) that constitute a net cost for him of \$3 but benefits his neighbor by \$6. His neighbor, in a similar situation, must make the same decision.

There is another implicit exchange beyond this, one which may be more frequent empirically, and that is not precisely equivalent to the others. If the two actors have a social relationship, consisting of a set of obligations and expectations (assumed for the present to be symmetric), various other exchanges are possible. If A_1 is to prevent an action of A_2, which imposes a cost on him of \$6, while A_2 benefits only by \$3 from the action, A_1 need only introduce into the negotiations some other event that he controls that has a cost for him of less than \$6 and a benefit for A_2 of more than \$3. A promise or a threat with respect to this event can serve A_1 perhaps as well as, perhaps better than, the action which is analogous to the action of A_2 which he wants to control. To state it differently, he need not use as a sanction for A_2 the same kind of action as the one he is sanctioning. If A_2 shows up late for a meeting, A_1 need not show up late for the next meeting; he can express disapproval, he can threaten to break off the meetings altogether, or he can offer A_2 a special benefit if he arrives on time for the next meeting. A_2 is in a similar position, so long as he has control over events that meet the

necessary criteria (of less cost to him than \$6, and of more benefit to A_1 than \$3).

Note how this changes matters. If A_1 has used an extrinsic event as a sanction for A_2, he has made no commitment on his original action. He remains free to contribute or not, or in the case of the meeting, free to show up late at the next meeting himself. But there is a second change as well: The other events may include some which have quite different internal and external effects than those of the original action at issue. In particular, the costs to A_1 of using one of these events as a sanction for A_2 may be very small, yet A_2's interest in the event may be sufficiently great that the sanction is effective.

It is important to recognize these additional sanctioning possibilities that actors may have for one another, because of the importance they lend to the existence of other events linking these actors. Attention to these additional possible sanctions is also important because of the potential sanctioning asymmetries that result from inequalities in their control of events of interest to one another.[9]

Beyond Two Actors

It is when pairwise exchanges (either isolated or in a market context) cannot bring about a social optimum that interests in a norm arise. This may be illustrated by expanding the joint project described earlier to a common project of three actors. Again, each has the alternative of contributing \$9 or nothing. For every \$3 contributed, the product will be \$4, an earning of \$1. The social product will be divided equally among the three.

Diagram 2.2 shows the outcomes for each combination of actions. Since the situation is symmetric for the three, the outcomes can be summarized more compactly, as shown in the tabulation given below.

	Gains or losses	
Number of contributions	Contributors	Noncontributors
0	—	0
1	−5	4
2	−1	8
3	3	—

Here, the situation is fundamentally different from before. It is no longer possible for two actors to exchange control over their actions and to gain by so doing. If there are no contributions, with a net gain or loss

		A_3: Contribute		A_3: Not contribute	
		A_2: Contribute	A_2: Not contribute	A_2: Contribute	A_2: Not contribute
A_1	Contribute	3, 3, 3	-1, 8, -1	-1, -1, 8	-5, 4, 4
	Not contribute	8, -1, -1	4, 4, -5	4, -5, 4	0, 0, 0

Diagram 2.2.

to each of \$0, and A_1 exchanges control with A_2, each contributing for the other, they end up losing \$1, with A_3 gaining \$8. If A_3 is already contributing, then A_1 and A_2 each gain \$4 before an exchange. If they exchange control, with each contributing for the other, the gain for each is \$3, making them \$1 worse off than without the exchange.

It is only if *both* A_2 and A_3 can be induced to change their actions from not contributing to contributing, contingent on A_1's contribution, that it becomes profitable for A_1 to join such an arrangement. In such a case, the outcome for each changes from a gain of \$0 to a gain of \$3. Thus a compact among the three is necessary to bring about a gain to each. One form of compact is a norm, with sanctions attached to enforce it. It is in this way that we can say that each comes to have interests in a norm.

The structure of interdependence in this case is one in which, if a norm arises at all, it will be a conjoint norm, with the same actors as both targets and beneficiaries. It will be an essential norm, not a conventional one, because there is one direction of action that benefits each (contributing), and one which does not. It would be possible to construct a similar artificial example and a similar matrix of outcomes with interdependence that generates interest in a conventional norm that is straightforward and self-evident and will not be presented here. For a disjoint norm, the matter is somewhat different; examination of such norms will be discussed later in the chapter.

Until this point, I have examined the conditions which lead to interest in the creation of a norm and the imposition of sanctions to bring about its observance. I have said nothing about the conditions which allow this interest to be realized by the bringing into being of a norm and sanctions. The question that must be answered is this: What is required to

get from interests in a norm to the actual existence of a norm backed by sanctions?

Social Structure and the Realization of Norms

The fundamental problem exhibited by the common project involving three actors in the example of the preceding section is a problem of social organization. In the two-actor project, each has the resources to prevent the other from imposing negative externalities upon him (or equivalently in this case, to induce the other to act in a way that brings positive externalities).[10] This is no longer so in the three-actor project. No single actor can exchange control with a single other to their mutual benefit. The externalities of the action of each for any one other actor are less than the actor's own effect on his gains. If a social optimum is to be achieved, something beyond pairwise exchange is necessary. One solution would be a sequence of pairwise exchanges in which first A_1 and A_2 exchanged rights of control, then A_2, with the right to control A_1's action, exchanged *this* for the right to control A_3's action. The rights of control are then distributed as follows:

A_3 controls a_1
A_2 controls a_3
A_1 controls a_2

If this pair of exchanges took place, then each would exercise the control he possessed in a way that benefitted him (as well as one of the other two): A_3 would commit A_1, A_2 would commit A_3, and A_1 would commit A_2. But the first exchange would take place only if both actors knew that a second exchange was possible—for without it, each gives up something worth more to him for something worth less. Furthermore, after the exchange between A_1 and A_2 has been made, A_3 finds it *not* to his benefit to exchange control with either. Thus the transactions would end after the A_1—A_2 trade, and both would end up losing \$1, while A_3 gained \$8.

Even if this were not so, the solution depends critically on a condition often not found: the knowledge that further transactions will be available to make an initially unprofitable exchange a profitable one. As is evident from the study of primitive systems of economic exchange, the development of such exchanges (in which objects come to have a "value" in exchange apart from their utility for the actor, leading the actor to acquire them for further exchange) is not a simple one (see Einzig, 1966).

There is one common device that is sometimes used by individuals

who anticipate benefits from a common activity to which all contribute, but have difficulty in overcoming the problem that each is better off by not contributing. This is to vest rights of control of their actions in a leader. This requires, of course, a high level of trust in the leader to act in terms of the followers' interests, trust which sometimes occurs when an actor is viewed as having charismatic qualities.

In the absence of such a "collective" solution to the public goods problem, some kind of combined action is necessary if a social optimum is to be obtained. This, in turn, depends on the existence of a relation between at least two of the three. As a first way of looking at this, Figure 2.1 shows two cases: In (a), actor A_1's action has an effect on A_2 and A_3 (as shown by the arrows) who have no social relationship with one another. Their social relations are with other actors A_4 and A_5. In (b), there are the same effects of A_1's actions, but actors A_2 and A_3 have a social relationship (the content of which I will discuss shortly). In figure 2.1(a), any sanction from A_2 or A_3 to direct A_1's action so it is not inimical to their interests must be applied by either independently; and in the three-actor common project the only actions available to A_2 and A_3 are the actions of contributing or not. Can either use that action as a sanction to bring about A_1's contribution? There are two obstacles to doing so.

First, there is a sequencing problem: If each acts independently and simultaneously, then A_2 and A_3 will already have acted when they discover A_1's noncontribution. The sequencing problem is too complex to go into here, except for two comments. First, since a sanction by A_2 and A_3 of not contributing cannot affect A_1's contribution in this project, it is necessary to consider an unlimited sequence of projects, with the sanction for the current project affecting A_1's contribution in the next project. Luce and Raiffa (1957) point this out for the case of a two-person

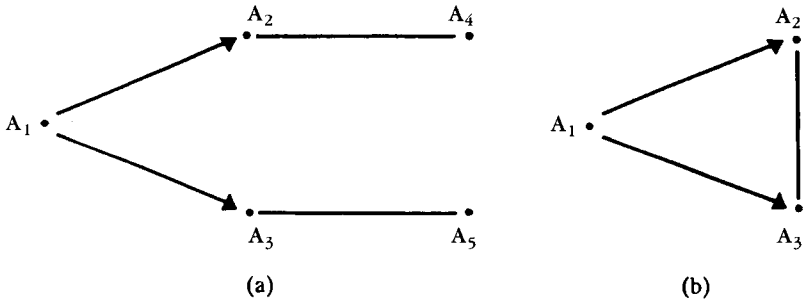

Figure 2.1. Two social structures within which an actor's (A_1's) actions have external effects.

prisoner's dilemma, having the same payoff structure as the common project. (The force of the assumption of an unlimited sequence is softened if the next project only occurs with probability less than 1.) In the subsequent discussion, I will ignore the sequencing problem for purposes of exposition, while recognizing that any effect of a sanction on a project has its effect on the next, and that the credibility of a sanction depends on a future in which there is an indefinite sequence of projects.

The second problem is that A_2's threat not to contribute is not a serious one: A_1 gains \$5 by not contributing, and A_2's sanction (not contributing) costs A_1 only \$4. Thus A_1 is still better off not to contribute, even if A_2 were to sanction by not contributing.

If, however, there is a social relationship between A_2 and A_3, as indicated in Figure 2.1(b), then, neglecting the problem of sequence, there is the possibility of joint action on the part of A_2 and A_3. A_2 and A_3 may jointly threaten not to contribute if A_1 does not contribute. This sanction would cost A_1 \$8, while the benefit of not contributing is only \$5. The threat is also a credible one, for A_2 and A_3 would gain \$1 by carrying it out. Thus apart from the sequencing problem, the relationship between A_2 and A_3 makes possible a joint sanction that would be effective for A_1. But what is meant by the unanalyzed term "social relationship"?

Unless it is possible to unpack the term "social relationship" and discover within it some interests and control that give one or both actors leverage vis-à-vis the other, the term can have no explanatory value in a theory based on rational choice. Because social relationships consist of obligations and expectations, either asymmetrically or symmetrically held, and because each actor continues to hold control of some events in which the other is interested, there exists inherently in each social relationship leverage to obtain a commitment from the other. If in Figure 2.1, actor A_2 has obligations toward A_3, then A_2 may pay off a portion of those obligations by sanctioning A_1. Or A_3 may initiate such an action by calling A_2's obligations. Or if A_2 has control of some event in which A_3 is interested (which may be nothing more than A_2's approval), A_2 and A_3 can carry out an implicit exchange, with A_3 incurring the cost of sanctioning A_1 in return for control of the event currently controlled by A_2. Note, however, that the discussion of a social relationship goes outside the payoff structure of Diagram 2.2. Introduction of "obligations" of A_2 toward A_3, or "some event" over which A_2 has control and in which A_3 is interested, introduces elements not shown in the payoff matrix.

This answer to the analytical question about how social relationships may be used to aid the employment of sanctions does not indicate how sanctions actually take place. I will state, in general, the problem of applying sanctions when neither can afford to do so separately, but will

not discuss the kinds of sanctions that actually are applied to enforce a norm, these are discussed in Coleman (1990, Chapter 11).

The Second-Order Public Goods Problem in the Enforcement of Norms

The sanctioning problem is what has been called the "second-order public goods problem," or the "second-order free-rider problem." To see just what is meant, it is useful to go back to one of Aesop's fables, *The Mice in Council*. The council meeting was called to discuss a problem faced by the mouse society: how to control the cat who was slowly decimating the population. In the terms used in this chapter, the cat's action imposed severe externalities upon the mice, and constituted, in effect, a public bad, creating constant danger for each. This is the first-order public goods (or in this case, public bad) problem.

The *second-order* public goods problem is indicated by the statement of the wise old mouse who finally rose in the council, after a proposed solution (that a bell be put around the cat's neck to warn of its approach) had been roundly applauded. He suggested that the council consider how the bell was to be fastened about the cat's neck, and who would undertake the task. The second-order public goods problem is the fact that, just as the cat's action imposes externalities upon all, an effective sanctioning of the cat's actions also has externalities (positive in this case) for all those experiencing the benefits of the sanction. Yet the benefits to the mouse who would undertake to bell the cat would not be sufficient to overcome the costs. Or in the case of the three-actor project shown in Diagram 2.2 the first-order public goods problem is the fact that each will benefit from the contribution of others; and the second-order public goods problem is the fact that if A_1 does not contribute, then the sanctioning of A_1 is a public good for A_2 and A_3. Neither receives sufficient benefits from his own action to motivate sanctioning A_1. The problem may not appear to be a serious one for the three-person common project. Just as sanctioning may be effected through an implicit or explicit exchange—without any notion of a norm—in a two-person common project like that in Diagram 2.1 the sanctioning public goods problem for one actor's failure to contribute to a three-actor common project is reduced to a two-actor common project. This may be solved whenever there is the possibility of exchange between the two actors who experience externalities from the third. One can compensate the other for the net costs of applying the sanction (that is, the costs, to A_2 say, of sanctioning A_1, less the benefits that A_2 will derive directly from the effects of the sanction). More generally, the second-order public goods problem of sanctioning is always one actor smaller than the original public goods problem.

It is now possible to state the second principle referred to at the beginning of this chapter, which concerns the conditions under which the demand for an effective norm will be satisfied. Stated simply, these are the conditions under which the second-order free-rider problem will be overcome by rational actors.

If there is a social relationship between actors, as between A_2 and A_3 in the example, then this overcomes the second-order free-rider problem for A_1's contribution. If there is a social relationship between A_1 and A_2, and between A_1 and A_3, which may or may not be strong enough to ensure A_1's contribution, but is strong enough to ensure A_1's joint sanctioning with A_2 or A_3, then the contributions of each are assured. The threat of sanctions exists, although they need not be applied.

The existence of a norm facilitates achievement of the social optimum by making use of the social relationships that exist in a social system to overcome the second-order free-rider problem. The existence of a norm implies that each has the right to sanction others for violating the norm, and the right to sanction others for not sanctioning those who violate the norm. It allows the introduction of other resources to be used in sanctioning, so that even when the conditions do not exist for a credible sanctioning threat within the resources of an activity like the common project shown in Diagram 2.2, the social optimum may be achieved through the use of other resources.

Heroic versus Incremental Sanctioning

In the examination of ways in which norms are characteristically enforced, it becomes clear that a common mode of sanctioning is one that can be characterized as incremental. This is exemplified in unions such as the typographical union by putting "in Coventry" scabs and others who seriously violate union norms (Lipset *et al.*, 1956). It is exemplified in communes like Bruderhof by sending a serious offender to an isolated dwelling and cutting off all communication (Zablocki, 1971). It is exemplified by the development of "reputations" followed by avoidance, or more generally what Merry (1984, p. 279) terms the third phase of gossip: some form of sanction such as snubbing, in which each participates as a result of the informal consensus achieved in the second phase of gossip. In incremental sanctioning, the cost incurred by each sanctioner is small, as are the effects, but they are additive, giving a total effect equivalent to that of a single large sanction.

Aesop's fable of the Council of mice, however, is a reminder that it is

not always possible to sanction incrementally. To bell the cat was not an activity that could be engaged in by additive increments. It required what I will call a heroic sanction, that is a sanction in which the total effect occurs through a single actor's sanctioning action. And in the examples of norms and sanctions given earlier in the chapter, the sanctions were imposed by single individuals.

In this section, I will use the example of the common project to examine the structure of action when sanctions are incremental, carried out by all the actors in the collectivity other than the one being sanctioned.

If sanctions can be additive in their effects, as the empirical evidence suggests they are in many cases, A_2 can bring about a contribution of half of the \$9 from A_1 through a sanction costing A_2 \$2 ½, and bringing benefits of \$2 each to A_2 and A_3. The net cost to A_2 is only \$½. This structure is shown in Diagram 2.3. Here there continues to be a prisoner's dilemma, but one with extensive possibilities for mutually beneficial arrangements, because of the disparity in the sanctioner's net loss (only \$½ and the other's gain (\$2) from the sanction. This example, however,

		A_3: Sanction	A_3: Not sanction
A_2	Sanction	.5, .5	-1.5, 1
	Not Sanction	1, -1.5	-1, -1

Diagram 2.3.

does not show the virtues of incremental sanctions as sharply as does a case with a larger number of actors. Consider the same common project, but now with six participants, each contributing 0 or \$9, with \$1 earned for every \$3 contributed, and the total product divided equally among the six. The net gain for each contributor and noncontributor in each configuration of contributions is shown in Diagram 2.4.

Here, the net loss incurred by contributing is no longer \$5, but \$7. (For

Number of contributions	Net gain (\$) for: Noncontributors	Contributors
6	—	3
5	10	1
4	8	−1
3	6	−3
2	4	−5
1	2	−7
0	0	—

Diagram 2.4

example, if five actors are contributing, the noncontributor's gain is \$10. If he contributes, he ends up with \$3, making him \$7 worse off.) The net gain experienced by the others from an actor's sanction is no longer \$4—it is only \$2. (For example, if the sixth actor does contribute, their net gain increases from \$1 to \$3.) If sanctioning cannot be incremental, the heroic sanctioner must incur a cost of \$7 to achieve a gain of only \$2. He has a net loss of \$5, rather than the \$1 for the heroic sanctioner in the three-actor project. Furthermore, this net loss of \$5 cannot be made up without loss by another who benefitted because that actor's benefit too from the heroic action is only \$2. Not even *two* others could provide sufficient rewards to the heroic sanctioner to make his action anything other than foolhardy. If they rewarded him with all their gains, he would still have a net loss of \$1. It would take three others, that is, all but one of the four who gained by his heroic sanction, to make his sanction no longer constitute a net loss.

If the sanctions can be incremental, the degree of exposure of the sanctioner is much less. A sanctioner, say A_2, would incur a cost of \$7/5, or \$1.4, and would gain from his sanctions alone \$0.4 from the incremental contribution made by A_1, the actor that is sanctioned (although he gains \$2 altogether from the total set of sanctions if others sanction as well). Thus each experiences a net loss of \$1 by sanctioning. As before, it would be possible for this to be made up by a sequence of rewards from others, of \$0.4 each, which would again require participation by at least three of the other four. Alternatively, additional incremental sanctions from the others would make up A_2's loss, each incremental sanction reducing A_2's loss by \$0.4. If all sanction incrementally, he gains \$0.6. Thus for incremental sanctions to pay the sanctioner, some prior collective decision that all (or at least many) will sanction (as in the "consensus" which Merry describes as the second stage of gossip) is required.

For an example, suppose all members of a club are expected to clean up after meetings, but one member consistently fails to help. If one person expresses disapproval, this might induce a small effort on the offender's part, but would also worsen the relation between these two, an effect that could be more important to the potential sanctioner than the benefit from the offender's efforts. But if all concurred in expressing disapproval, inducing the offender to make his full contribution, the benefits to each would outweigh the costs of the worsened relation with the offender.[11]

Returning to the example of the common project, suppose there were not a binding collective decision, and all but one sanctioned. Then the sanctions could go one stage deeper. A_1 is the noncontributor, and suppose that A_2 is the nonincremental sanctioner. Each of the others has provided an incremental sanction, and A_1 has made four fifths of his total contribution. A_2, who is \$1 better off by not sanctioning, can be induced to sanction again either by a heroic second-stage sanction of \$1, which works out to a net cost of \$0.6 to the second-stage sanctioner, or by incremental sanctions of \$0.25, which work out to a net cost of \$0.15 for each of the sanctioners.

The overall difference between the heroic and the incremental sanctions lies in the magnitude of sanction required at every stage. At the first stage, in the six-actor project, the heroic sanctioner must incur a loss five times that of each incremental sanctioner. In this project, A_2, the heroic sanctioner, incurs a cost of \$5. At the second stage, for A_3 to reward the heroic sanctioner heroically imposes a net loss of \$3 on A_3.

If sanctioning is incremental, the free-rider problem remains, but at a greatly reduced magnitude. The net cost to each sanctioner is \$1, rather than \$5. If the second-stage sanction (the reward to the incremental sanctioner) is heroic, that means only a net cost of \$0.6 to the second-stage heroic sanctioner, rather than \$3. If the second-stage sanction is also incremental, the net cost to each of the four sanctioners is only \$0.15.

What this means in practice is that in many circumstances in which heroic sanctions are beyond the resources of any sanctioner, they are readily available for incremental sanctioning. These resources may consist of other events which are controlled by each of the potential sanctioners. The values specified above indicate only the maximum costs that the sanction can impose on the sanctioner, and the value it must have to the sanctionee. When the sanctions are of such small cost as they come to be in a large group, a positive sanction may consist of nothing more than a credit slip in the form of gratitude for what the other has done, or a negative sanction may consist of nothing more than a withdrawal of credit in the form of displeasure ("Wait till you ask me to do something for you.")

Other possibilities exist as well with incremental sanctioning. If there is some heterogeneity among the potential sanctioners the free-rider problem may be overcome at some stage, and in any case will constitute a lesser obstacle. The complex possibilities that exist can only be alluded to here.

There is one additional point: the use of the term "heroic" here refers to a single sanction by a single sanctioner of sufficient size to bring about A_1's contribution. If the set of five contributors, or a large enough subset, can act as a single actor, there can be a single sanction from that set sufficient to bring about the contribution and yet bringing a net benefit to each. The frequent institution, in communes, of meetings once a week or at another regular interval, at which the whole membership gathers to hear self-criticism or criticism by others suggests that in such settings this method of sanctioning is easier to organize than either heroic or independent incremental sanctioning.

Conclusion

I have attempted to show that two principles can account for emergence of norms with effective sanctions. One of those is a principle concerning the conditions under which a demand for norms will arise. The conditions involve the existence of similar externalities from a focal action for a set of potential holders of the norm.

The second principle concerns the conditions under which the demand will be satisfied by effective sanctions. These conditions involve the potential for sanctions internal to the set of norm holders, to overcome the "second-order free-rider problem."

Notes

1. See O'Flaherty and Derrett (1978), and Kunst (1978).
2. Ullmann-Margalit (1977:97) calls these coordination norms, and distinguishes between those that arise through convention and those adopted by decree. I will not make use of this distinction.
3. Ullmann-Margalit distinguishes three kinds of norms, which she calls "prisoner's dilemma norms" "coordination norms," and "norms of partiality." These correspond approximately to what I have termed essential norms, conventional norms, and disjoint norms, respectively. However, the correspondence is not complete, because essential norms, to use my terminology, may be disjoint or conjoint, while Ullmann-Margalit's three classes are mutually exclusive.

4. It is, of course, the case that when academic activities are organized interscholastically, they too can generate a prescriptive norm. Striking cases of this may be found in a description of rural schools in Kentucky engaging in statewide competition in academic subjects (Stuart, 1949, p. 90ff).

5. A game with payoffs showing this structure is called a prisoner's dilemma. See Luce and Raiffa (1957) or Rapaport and Chammah (1965) for a discussion of this game.

6. So far as I know, Erling Schild and Gudmund Hernes were the first to (independently, in 1971) point out that the simplest social solution to the prisoner's dilemma is exchange of control between the two players, an action which is rational for each. Peter Bernholz (1987) has shown that the Sen paradox of a paretian liberal, where the payoff structure is that of a prisoner's dilemma, is solved in the same way.

7. This is a fundamental point on which Axelrod (1984:49), who discusses the growth of cooperation in two-person prisoner's dilemmas, exhibits confusion. At some points he can be interpreted as asserting that pairwise interactions in large populations where the same two parties meet only very infrequently, will generate the same cooperation as found in his pairwise "tournaments." In general, however, Axelrod's work in that book demonstrates the point made here: that bilateral exchanges, explicit or implicit, are sufficient, without introduction of a norm, to arrive at a social optimum in pairwise interactions with externalities. See Coleman (1986) for examination of social structural conditions in which the contact does not allow such agreements, implicit or explicit, to be effective.

8. It is worth noting that in Ullmann-Margalit's two-person example (two mortarmen in isolated outposts) in which a norm of "honor" is seen as one solution where a bilateral exchange cannot be, the latter is precluded because no prior agreement can be carried out after the fact since in her example, one mortarman is dead. It is also true that the norm of "honor" arises more broadly in military units, where one soldier may be risking his life save the lives of a number of his fellow soldiers, not merely one.

9. This exposes also another source of asymmetry, hidden by the symmetry of the example. Even for activities in which all actors' similar actions impose externalities on one another, the externalities may be unequal, providing sanctioning opportunities for some actors that do not exist for others. This is related to questions of interpersonal comparison of utilities, and as will be evident later, a correct untangling of that issue will be important for the analysis of norms as for other aspects of the social system.

10. As indicated earlier, when there are only two alternative actions, as in this case, there is no distinction between prescriptive and proscriptive norms.

11. Empirically the costs might also be reduced, for disapproval from all might lead the offender to accept the collective verdict, and not respond unpleasantly to the members expressing disapproval. However, in the example of Diagrams 2.2 and 2.3, the net gain from incremental sanctioning by all does not depend on such reduced costs.

References

Axelrod, Robert. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Bernholz, P. (1987). "A general constitutional possibility theorem." Pp. 383–400 in G. Radnitzky and P. Bernholz, eds., *Economic Imperialism*. New York: Paragon House.

Coase, Ronald H. (1960). "The problem of social cost." *Journal of Law and Economics* V(3):1–44.

Coleman, James S. (1986). "Social structure and the emergence of norms among rational actors." In *Paradoxical Effects of Social Behavior: Essays in Honor of Anatol Rapoport*. eds. A. Diekmann and P. Mitter. Pp 55–83. Vienna: Physica-Verlag.

Coleman, James S. (1987). "Norms as social capital." Pp. 133–155 in G. Radnitzky and R. Bernholz, eds., *Economic Imperialism*. New York: Paragon House.

Coleman, James S. (1990). *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.

Dahrendorf, Ralf. (1968). *Essays in the Theory of Society*. Stanford: Stanford University Press.

Einzig, Paul. (1966). *Primitive Money*, 2nd ed. London.

Garnsey, Peter. (1973). "Legal privilege in the Roman Empire." Pp. 146–166 in Donald Black and Maureen Mileski, eds., *The Social Organization of Law*. New York: Seminar Press.

Kunst, Arnold. (1978). "Use and misuse of Dharma." Pp. 3–17 in O'Flaherty, W. D., and J. D. M. Derritt, eds., *The Concept of Duty in South Asia*. New Delhi: Vikas.

Lipset, S. M., M. A. Trow, and J. S. Coleman. (1956). *Union Democracy*. New York: Free Press.

Luce, R. D., and Howard Raiffa. (1957). *Games and Decisions*. New York: John Wiley and Sons.

Merry, Sally E. (1984). "Rethinking Gossip and Scandal." Pp. 271–302 in D. Black, ed., *Toward a General Theory of Social Control, Vol 1*. New York; Academic Press.

O'Flaherty, W. D. and J. D. M. Derrett (eds.). (1978). *The Concept of Duty in South Asia*, p. xiv. New Delhi: Vikas.

Rapaport, A., and A. M. Chammah. (1965). *Prisoner's Dilemma*. Ann Arbor: University of Michigan Press.

Sorokin, P. (1928). *Contemporary Sociological Theories*. New York: Harper and Row.

Stuart, J. (1949). *The Thread That Runs So True*. New York: Charles Scribner.

Ullmann-Margalit, Edna. (1977). *The Emergence of Norms*. Oxford: Clarendon Press.

Zablocki, B. (1971). *The Joyful Community*. Baltimore, Maryland: Penguin Books.