2 A THEORY OF THE STATE AND OF SOCIAL ORDER

SUN-KI CHAI and MICHAEL HECHTER

University of Arizona

After a long period of neglect, the problem of social order has recently revived in social science discourse. This revival is largely due to the efforts of theorists attempting to explain the conditions under which rational egoists cooperate. One line of analysis offers invisible hand theories suggesting how cooperation may emerge spontaneously among rational egoists under anarchy (Axelrod, 1984; Ullman-Margalit, 1977; Taylor, 1987; Calvert, 1991, 1994). Another, drawn from the new institutional economics, focuses on the state's role in creating and maintaining cooperative equilibrium between individual actors (North, 1981; 1983; 1990; North and Weingast, 1989; see also Hardin, 1991).

Although these two kinds of theories have proven insightful, neither can account for the emergence of order in large groups, let alone national societies. The former theories assume that the attainment of order in national societies is fundamentally similar to that in small groups. However, it is implausible that large numbers of atomized individuals can cooperate as the members of small groups do: the coordination and monitoring problems are simply too severe (Hechter, 1993: 2). The latter theories simply take the state's existence for granted. Yet if – as is likely – the state must be invoked to solve the problem of order, then its formation and maintenance should be part of the explanation.

If atomized individuals cannot produce social order on their own, one alternative is to assume that cooperation can take place at a higher level of analysis – among a few large groups that encompass many individuals. The apparently insuperable difficulty involved in coordinating the actions of many disconnected individuals then reduces to the more manageable problem of coordinating the actions of a
Part I offers a formal, rational choice model of group solidarity. It analyzes how the incentives for free riding affect the cost of attaining group solidarity through enforcement of member compliance with obligations to contribute to a set of joint goods. Next we turn to dependence - in particular, the opportunity cost to members of exiting from a group. Dependence affects the cost of enforcement and determines the upper bound on the size of contributions to joint goods that can be maintained without having the group unravel. Finally, we discuss coercion, the forcible removal of the exit option. Although high levels of dependence are conducive to group solidarity, under certain conditions they also lead to coercive control by one subset of members over another.

Part II extends this model to explain the formation and maintenance of the state. It proposes that states maintain order between groups in a fashion isomorphic to that which group control institutions maintain solidarity among individuals. A high level of solidarity is a prerequisite to the consideration of groups as unitary actors at the state level, and thus to their capacity to contribute to social order. Hence, the state emerges either as the control institution of a supergroup created by a small number of highly solidarity groups to prevent mutual predation, or as a means used by one set of groups to impose its will on others.

The arrangement of these groups has important implications for social order (Part III). Whereas some form of group mediation between individuals and the state is always necessary to minimize the costs of maintaining social order, the optimal configuration of groups varies across social forms. Sometimes complex structures made up of multiple layers of nested groups are optimal for attaining social order; at other times simpler structures will do the trick. The optimal configuration of groups for the attainment of social order is a function of environmental influences on monitoring costs. These, in turn, affect the optimal size of groups. Dependence affects both the ability of groups to contribute to social order and the coerciveness of the resulting order. A brief summary of the argument and its implications for further research concludes.

PART I: A MODEL OF GROUP SOLIDARITY
We view groups not merely as collections of rational individuals, but also as encompassing control institutions shaping interactions between members. For analytic purposes, our initial focus is on voluntary groups whose members are free to exit and yet who choose to remain.
If individuals are rational actors (that is, maximizers of expected utility), then they will form a group only if membership provides each with a higher level of expected net benefit than that attainable elsewhere.\(^4\) Hence the original impetus for group formation is the prospective member’s desire to obtain the group’s joint goods. Absent institutionalized and enforced mutual agreements, individuals are likely to engage in mutual predation, even if this results in lower net benefits for all concerned.

To consume the joint goods that motivate membership, individuals must agree to a set of corporate obligations (or rules) that permit the goods in question to be produced. However, the mere formation of a group and agreement to a set of corporate obligations (including, but not limited to production rules) is insufficient to guarantee compliance to them. Even if all the members have agreed to honor these obligations, each will do better by ignoring them when doing so is expedient.

Because ensuring compliance with corporate obligations is necessary to gain access to valued joint goods, members have an interest in creating a group-level control institution for monitoring and sanctioning one another. This institution perforce constrains members’ freedom of action to some degree. By deterring free riding, the control institution assures actual and potential members that their contributions to the group will redound to their personal benefit. Group solidarity will be defined as the extent to which members comply with their corporate obligations to contribute to the group’s joint goods (Hechter, 1987).

Let \( n_s \) be defined as the number of members in group \( s \) and \( c_{yw} \) as the cost borne by member \( y \) when she complies with her corporate obligation to contribute to the group’s joint goods. In essence, \( c_{yw} \) is the gap between the net benefit a member would receive from compliance with this obligation, and the net benefit she would receive were she to free ride with impunity.\(^5\) We do not attempt to determine the exact level of contributions required to produce a particular group’s joint goods; this is left exogenous to the model. For the sake of simplicity, we initially assume that members are faced with one of two alternatives: to comply and absorb the associated costs, or to fail to comply and face sanctions if detected. A group will be assumed to punish known free riders with a one-time set of sanctions, after which these members can continue to participate with no further punishment, barring future noncompliance.

Therefore, being rational actors, members will comply only if the probability of detection multiplied by the sanctions imposed given detection equals or exceeds the benefits from noncompliance. Hence, a condition for compliance of member \( y \) in group \( s \) is \( p_{yw} d_{yw} \geq c_{yw} \), where \( p_{yw} \) is the exogenously determined probability of detecting noncompliance in group \( s \) by member \( y \), and \( d_{yw} \) is the amount of punishment meted out to member \( y \) in the event of detection. Note that the benefits an individual gains from group membership do not enter directly into this particular calculation, since they will be enjoyed regardless of compliance unless withholding them is part of the punishment.\(^6\)

Assume now that \( p_{yw} \) is a function of the size of the group \( n_s \), is equal for all individuals in the group (and hence can be referred to as \( p_s \)), and is differentiable and monotonically decreasing, that is \( \partial p_s / \partial n_s < 0 \). This assumption is justified by the intuitive notion that visibility is decreased in large groups, resulting in greater difficulty in monitoring. The exact shape of \( p_s \) may vary depending on a number of additional factors relating to characteristics of the surrounding environment.\(^7\)

Sanctioning will be carried out by a control institution, which is in turn funded by group resources. Hence resources above and beyond those necessary to provide the joint goods must be expended by the group to generate and maintain its sanctioning capabilities.\(^8\) These resources can quite reasonably be assumed to be a monotonically increasing function of the sum of sanctions \( d_{s1} + \cdots + d_{sn} \) that must be imposed on each noncompliant member. Since it is optimal, if sanctions are costly, to provide just enough sanctioning to yield compliance, the amount of sanctions imposed will for detected noncompliance equal the cost of compliance divided by the probability of detection. This implies that the amount of sanctions increases as the probability of detection falls, and therefore that they increase with group size, since the probability of detection is inversely related to group size. Formally, for each member \( y \), \( d_{yw} \) will be equal to \( c_{yw} / p_s \). Hence, it follows that \( \partial d_{yw} / \partial p_s < 0 \). Since \( \partial p_s / \partial n_s < 0 \), it follows that \( \partial d_{yw} / \partial n_s > 0 \), since \( n_s \) has no effect on \( d_{yw} \) other than through \( p_s \).

Let \( \gamma_{yw} \) indicate the amount of resources needed to maintain minimal adequate sanctions to ensure contributions from all members of the group \( s \). The superscript indicates “first-order” resources, to distinguish it from other levels to be introduced later. It can reasonably be assumed that \( \gamma_{yw} \) is equal to some differentiable function \( \Phi \) of \( d_{yw} \) such that \( \Phi > 0 \) and \( \Phi(0) = 0 \), where \( d_{yw} \) is defined as \( d_{s1} + d_{s2} + \cdots + d_{sn} \). It follows that \( c_{yw} / \partial n_s > 0 \) and \( \gamma_{yw} / \partial c_{yw} > 0 \) for all \( y \). \( \gamma_{yw} \) represents a cost of control in group \( s \) that must be borne by the group’s membership. Hence, the existence of a control institution creates
additional obligations over and above those entailed in merely producing the joint goods, and the extensiveness of these additional obligations is a monotonically increasing function of the original ones. These additional obligations – to fund the control institution – can be considered to be second order. Since second-order compliance entails additional costs to members, this increases the amount of sanctions that must be imposed on members who fail to comply with their obligations, and the resources for these additional sanctions must be provided by the membership. This creates third-order obligations, which in turn entail still more costs, and so on indefinitely. In other words, costly obligations have a multiplier effect because they increase the cost of control institutions in a recursive manner.

Let \( \gamma^j_z \) stand for the aggregate \( j \)-th order costs for the control institution of the group – that is, the resources required for monitoring and sanctioning first- to \( j \)-th order compliant behavior. Let \( \gamma^0_y \) be the portion of those resources provided by member \( y \). By definition, \( \gamma^{j+1}_z = \gamma^j_z + \gamma^{j-1}_z + \cdots + \gamma^1_z \). Under these conditions, the aggregate amount of punishment that needs to be provided to ensure that member \( y \) engages in \( j \)-th order compliance is \( \gamma^j_y / p_y \) for \( j > 1 \). If so, then \( \gamma^j_y = \Phi((c_y + \gamma^{j-1}_y) / p_y) \) where \( c_y \) is defined as \( c_{y1} + c_{y2} + \cdots + c_{yn} \). The aggregate cost of the group control mechanism is \( \gamma^0_y \), which for brevity can be called \( \gamma_y \). Its value is \( \lim_{j \to \infty} \Phi((c_y + \gamma^{j-1}_y) / p_y) \). Provided that \( \gamma^j_z \) converges, this implies that \( \gamma_y = \Phi((c_y + \gamma_y) / p_y) \). The value of \( \gamma_y \) in terms of \( c_y \) and \( p_y \) will depend on the precise form of \( \Phi \). If \( \gamma^j_y \) does not converge, then clearly \( \gamma_y \) will have no finite value, and it will be impossible to ensure compliance with group obligations. It follows that the condition for solidarity in a group is that, given a distribution of contributions \( (c_{y1}, c_{y2}, \ldots, c_{yn}) \), there exists a distribution of second and higher-order contributions for enforcement \( (\gamma_1, \gamma_2, \ldots, \gamma_y) \) and a distribution of punishments \( (d_{y1}, d_{y2}, \ldots, d_{yn}) \) such that \( \forall y: p_y d_{yj} \geq c_{yj} + \gamma_y \) and \( \gamma_y = \Phi((\Sigma_{y1} + \Sigma_{y2} + \cdots + \Sigma_{yn}) / p_y) \), where \( \gamma_y \) is shorthand for \( \gamma^0_y \).

One major conclusion of this analysis is that the creation and maintenance of a control institution raises the extensiveness (hence, cost) of corporate obligations, and, through a multiplier effect, does so by an amount that is positively related to the size of the initial obligations. Another conclusion, however, is that by causing shifts in \( p_y \), and hence in the magnitude of the multiplier effect – changes in group size can lead to major shifts in the total sanctioning costs.

These two conclusions indicate that the extensiveness of (first-order) group obligations and group size interact in raising the cost of group solidarity. This helps explain why groups demanding particularly extensive obligations (including cults, communes, underground revolutionary cells) tend to be relatively small. Even with modern technology, it is extremely difficult to enforce extensive obligations in large groups. This becomes evident when we discuss the forces that bind members to groups, namely dependence and coercion.

Dependence and the Boundaries of Groups

To this point in the analysis, attaining group solidarity has involved the creation of a control institution sufficiently effective to ensure compliance with corporate obligations. As long as there is some level of sanctions at which the aggregate \( j \)-th order costs do not exceed the level of sanctions multiplied by the probability of detection, group solidarity can be maintained.

This seems to indicate that group solidarity can be maintained no matter how extensive corporate obligations are, if total sanctions for first- and higher-order noncompliance can be set high enough: \( \gamma^j_y \) converges to a finite number. Otherwise, there is no limit on the extensiveness of the obligations a group can demand of its members, even if the resulting marginal benefits do not outweigh the costs.

However, to this point the analysis has ignored the benefits that individuals gain from membership in the group. More specifically, it is clear that members will not voluntarily join a group unless the benefits to be gained from membership exceed the costs. Let \( b_{sy} \) stand for the benefits gained by member \( y \) from the provision of the joint goods of group \( s \). In order for a voluntary group to be viable, not only must enforcement costs \( \gamma^j_y \) converge, but it must also be the case that joint benefits exceed aggregate first- and higher-order costs, that is, \( b_{sy} > c_{sy} + \gamma_{sy} \) for all \( y \). If this is not so, then even if it is possible to induce compliance in equilibrium, such an equilibrium will never be reached because members will refuse to form a group.

Yet the analysis needs to pay heed not only to the absolute level of net benefits gained from membership in a group, but also to the difference between these net benefits and those that can be gained outside the group. This brings up an important characteristic of groups, namely the dependence of their members. Let the dependence of group member \( y \) on group \( s \) be defined as the opportunity cost of exit from group \( s \). This is the difference between the net benefits gained from membership in the group – that is, benefits \( b \) minus the aggregate costs \( c_{sy} + \gamma_{sy} \) – and the net benefits (the “reservation wage”) that the individual can gain outside a group or as a member
of some alternative group. So far we have only analyzed the relationship between the costs of compliance, the size of different groups and the cost of maintaining order within different groups, implicitly assuming that benefits were uniform across different groups.

The assumption of uniform benefits may not be justified, however. Some groups may be able to provide greater benefits than others relative to costs of first-order compliance $c_{y}$ and costs of monitoring and sanctioning $\gamma_{y}$. Individuals will realize greater benefits to the extent that a particular joint good is widely valued by members; then the group can specialize in producing that good. Contributions for the production of goods that are valued only by a subset of members do not provide any benefit for the rest.

Another type of dependence can be created by shared altruism among group members. To the degree that members are altruistic, then they benefit from each other’s consumption of joint goods, in addition to that derived from their own consumption. Formally, altruism can be regarded as a shared preference for the joint good that is comprised of increasing members’ welfare (Chai and Wildavsky, 1994; Chai, forthcoming). These benefits must be foregone when a person exits a group.

Dependence may also be created by complementary skills and capabilities. These may include similar skills such as a common expertise in the production of a mutually valued good, or commonly shared language and other modes of communication. Conversely, they may consist of a distribution of different skills among members that are necessary for the production of a particular good. A final source of dependence, one that favors the status quo, derives from the costs of moving from one group to another.

Typically, dependence is strongest among groups with a long history of interaction. Members of long-lived groups have had the opportunity to develop friendships and shared preferences that bind people together. Long-lived groups therefore are more effective at providing joint benefits given equivalent size and obligations than those formed more recently. This chapter will not attempt to analyze the differential consequences of various sorts of dependence. Instead, it will look at the effects of dependence in general on the maintenance of group solidarity.

Suppose $b_{y}$ is the joint benefit provided by group $s$ for member $y$ during each period. If $y$ complies, then $y$’s net benefits for membership in the group will be $b_{y} - c_{y} - \gamma_{y}$. Then the expected loss $y$ suffers per period on leaving $y$ and joining another group will be $b_{y} - c_{y} - \gamma_{y} - \max(t \neq s: b_{s} - c_{s} - \gamma_{s})$. In that case, the dependence of $y$ on $s$ will be based on this amount, presumably extended over an arbitrarily large number of future periods (for simplicity, we will not consider the possibility that $y$ may rejoin $s$ at some future period). If $z_{y}$ refers to this dependence, then $z_{y} = (1 + \delta + \delta^{2} + \cdots) (b_{y} - c_{y} - \gamma_{y} - \max(t \neq s: b_{s} - c_{s} - \gamma_{s})) = (b_{y} - c_{y} - \gamma_{y} - \max(t \neq s: b_{s} - c_{s} - \gamma_{s})) / (1 - \delta)$, where $0 < \delta < 1$ is the discounting factor for the benefit that $y$ gains in future periods.

If they are rational actors, prospective members will not voluntarily agree to join a particular group unless the benefits that can be gained in that group are larger than those attainable in all other alternative groups, that is, $z_{y} > 0$. This is in effect a more stringent condition than the condition that $b_{y} > c_{y} + \gamma_{y}$, i.e., $b_{y} - c_{y} - \gamma_{y} > 0$, since it is always possible for a member to eschew all groups (which is equivalent to forming a “group” consisting only of itself), in which case its benefits and costs $b_{y}$, $c_{y}$, and $\gamma_{y}$ will all be equal to 0.

As noted above, a member will comply only if the expected punishment for noncompliance $p_{d}$ is greater than the cost of compliance $c_{y} + \gamma_{y}$. However, note that the dependence of member $y$ on voluntary group $s$ will be positive. Then another option is available besides forgivable punishment, and that is the expulsion of the member from the group if the member fails to comply and is detected. The expulsion option is attractive in that it does not require the maintenance of a sanctioning force; thus it creates incentives for compliance to production rules without additional costs in the form of second- and higher-order contributions.13

Hence expulsion can be used, either alone or in tandem with forgivable punishment, in order to ensure contributions to the joint good. The functional form of the second- and higher-order costs of maintaining order will then be $\gamma_{y} = \Phi(e_{y} + \gamma_{y} - z_{y} - \gamma_{y})$, where $z_{y}$ is defined as $z_{y} + z_{y+1} + \cdots + z_{y+n}$. Hence $\gamma_{y} = \Phi(e_{y} + \gamma_{y} - z_{y})/p_{y}$ when $c_{y} = \gamma_{y}$ for all $y$. If there exist $y$ for which $c < z_{y}$, then we have $\gamma_{y} = \Phi(e_{y}/p_{y})$, where $e_{y} = c_{y} + \gamma_{y} - z_{y}$ when $c_{y} + \gamma_{y} - z_{y} > 0$ and $e_{y} = 0$ otherwise. If $c < z_{y}$ for all $y$, then $\gamma_{y} = \Phi(0) = 0$; thus there will be no second- or higher-order costs.

As this implies, dependence lowers the cost of maintaining solidarity in a group by reducing the extent to which sanctions need to be used to ensure compliance with group obligations. In doing so, it increases the level of group obligations (needed to obtain the joint goods) that can be required of members within a group, while maintaining sufficient incentives for members to join and remain in a group.

Dependence therefore puts a ceiling on the extensiveness of obligations that groups can demand of their members. It follows that groups
with high levels of member dependence can demand more extensive contributions than those whose members are less dependent. Clearly, more extensive obligations sometimes can more than compensate for their costs due to the greater joint goods they provide. Thus members actually may gain additional benefit from increasing their dependence, thereby making their exit option more remote. Paradoxically, the reduction of outside options for group members can increase the absolute benefit they can gain by remaining in their group.

Coercion and its Relationship to Dependence

Thus far we have only considered groups in which membership is entirely voluntary. However, in the absence or shortage of dependence among some subset of group members, one other option for preventing wholesale exit from a group exists—coercion. Groups may affect members’ calculus of compliance by restricting their exit options rather than by changing their net benefits for remaining in the group. More specifically, a coercive group control institution may punish those who attempt to leave, and therefore may be able to demand more extensive obligations from certain members than a voluntary one.

Coercion, however, requires the expenditure of additional sanctioning resources beyond those necessary to ensure compliance with corporate obligations in voluntary groups. This can be modeled by considering a given member’s incentives to exit as a kind of negative dependence. If a subset of members feels that they can gain more outside a group than within it, they will ride and simultaneously attempt to exit the group unless the expected amount of punishment exceeds both the gains from free riding and the gains from exit. This implies that the amount of resources required to coercively keep a member from attempting to exit is \( \Phi((c_x + \gamma - z) / p) \), where in this case, \( z < 0 \).

The costs of the coercive apparatus designed to keep certain members in a group must be borne by the membership or some subset of it. This subset may be the coerced members themselves, who thus may be required to support the very structure keeping them trapped within a suboptimal group. This kind of system may be feasible as long as the marginal costs of maintaining sanctioning resources are sufficiently low (otherwise the feedback costs of second-order rule-abiding behavior created by the need to fund the coercive apparatus will create additional incentives to exit that more than offset the deterrent effect of the apparatus), and as long as coerced members form only a small portion of the group’s population.

If coerced members contribute a large percentage of the resources expended on coercive controls, however, then the possibility of simultaneous coordinated exit threatens the entire edifice. If one member attempts to leave the group, that member’s contribution to the coercive apparatus is foregone, at least during the period when the exit is attempted. This does not significantly affect the viability of the coercive apparatus. However, if a large number of members attempts to exit simultaneously, their failure to contribute during this period will render the coercive apparatus so weak that it may be unable to prevent their exit.

For this reason, coercion affecting more than a small proportion of the membership of a group often will be based on the contributions of those who are not being coerced (this in turn implies differences among members in their levels of dependence). The additional demands placed on dependent members to support the coercive apparatus require higher monitoring and sanctioning costs, which will have recursive effects that raise the cost of maintaining solidarity in the group. These additional burdens may give some previously uncoerced members an incentive to free ride and exit if they are found out. This leads to a feedback effect—an even greater number of members will have to be coerced into staying in the group. Hence the resources that have to be devoted to coercion will increase apace.

If nondependent members comprising some large proportion of the group are to be coerced from exiting the group, then some core of voluntary members must tax themselves to contribute toward the coercive control institution. This core must be highly dependent, else its members will exit rather than bear the cost of the coercive control institution. Hence the only groups that can coercively restrict the exit of large proportions of discontented members are those with other members whose dependence is high enough (superdependence) to offset the (additional) costs of coercion. This analysis therefore shows how the factors responsible for group solidarity can also facilitate coercion in a group.

It may seem counterintuitive to say that dependent members of a group would ever be in a position to coerce nondependent members, since nondependent members, with an available exit option, would have greater power in any bargaining that takes place between them. However, this can be understood by examining the relationship between bargaining power and incentives for coercion. We define
coerced membership as a condition in which a member is forced to stay within a group even though she could do better elsewhere. Hence nondependence is a precondition for coerced membership. Those who are dependent do not need to be coerced, because they are willing to stay within the group voluntarily.\textsuperscript{18} It is precisely because they lack bargaining power that dependent members may be willing to resort to coercion of nondependent members rather than negotiation. We are discussing incentives for coercion, and do not imply that the dependent members will always have the resources to carry out coercion. Nonetheless, dependent members will better be able to ensure internal solidarity among themselves than nondependent members. Hence they may more credibly threaten to collectively use sanctions to enforce prevent nondependent members from leaving, even if the provision of these sanctions involve costs.

PART II: GROUP SOLIDARITY, THE STATE AND SOCIAL ORDER

Part I described the conditions for the creation and maintenance of solidarity in relatively small groups. Why does this analysis have any implications for the attainment of social order? Evidently, group solidarity is not a sufficient condition for social order. On the one hand, control costs increase dramatically with group size. On the other, there is no guarantee that a number of highly solidarity groups coexisting in the same environment will interact peacefully. Instead, some of these groups may promote disorder by using force to appropriate resources from more vulnerable ones.\textsuperscript{19} To interact cooperatively, these groups must remain solidarity and at the same time establish an institution that binds them together externally.

By now, this problem should be familiar. We proceed to extend the previous analysis to include supergroups whose individual members are not persons but groups. Note that this upward shift in the level of analysis does not fundamentally change the mechanism of group solidarity. By what sleight of hand can we possibly consider groups to be unitary actors? By no sleight of hand: the higher the solidarity of a group, the more it can behave as a unitary actor. To the degree that groups can control their own members, they can be considered as unitary members of higher-level groups. The rationale behind this assumption is straightforward: if groups can maintain internal solidarity, then they can guarantee that their members will not violate obligations enacted at a higher level of aggregation. If groups are solely responsible for monitoring and sanctioning their members, higher-level control institutions need not deal directly with these members.\textsuperscript{20}

Thus initial state can be seen as the control institution of a supergroup (society) encompassing the various highly solidarity groups of a given territory. For our purposes, societies differ from solidarity groups in that they are much larger in scale. States are formed for the self-same reasons that group control institutions are formed – either as a voluntary agreement among groups, or as a coercive structure used by one set of groups against another. In the voluntary case, a state emerges in an anarchical environment when a number of groups come together to form a single control institution that resolves intergroup conflicts and reduces their predation. Up to a point, therefore, we follow the basic Hobbesian logic on the formation of the state – with the important proviso that in this part of the theory highly solidarity groups are the relevant units of analysis, rather than individuals. Just as individuals who may have a common interest in a good are unlikely to obtain it without a group control institution, so highly solidarity groups with a common interest in social order are unlikely to obtain it without the state.\textsuperscript{21} In the coercive case, a state is used by a coalition of groups to harness its combined force to dominate another set of groups. However, high group solidarity is a precondition for the attainment of social order via the formation of a state, be it voluntary or coercive.

Group solidarity can occur at more than one layer below the level of the state. After all, the members of groups can themselves be groups. The state can be regarded as the top level of a hierarchy of nested groups control institutions extending down to the individual level.\textsuperscript{22} The attainment of order can be viewed analogously at each level of nesting, with the unit that is macro on one level becoming micro on the next highest level (Alexander et al., 1987). Group members that are themselves groups will have to expend resources sufficient to maintain internal order for them to be treated as virtual unitary actors, or corporate individuals (Coleman, 1990; 1993), at the next higher level of analysis. This will hold true down to the atomic levels where group members are single individuals.\textsuperscript{23} This analysis does not imply that individual rationality in groups always leads groups to act in a collectively rational manner. Such a claim is patently false (Hannan, 1992). Instead, the model specifies a set of necessary conditions (those promoting group solidarity) that are required for a group
of individuals (or lower-level) groups to behave as virtual unitary actors. There are a number of ways in which a group can be considered to be a rational actor. The most restrictive is when the group maximizes the utility function of each of its individual members. This generally is impossible unless everyone in the group shares exactly the same utility function. In our model, solidary groups are seen as rational because they behave as if they were corporate actors having a single utility function. However, this utility function need not necessarily represent those of all its individual members (Coleman, 1990, Part III).

Hence, attaining solidarity in a group is analytically similar to attaining order in a territory composed of multiple solidary groups; both are instances of the same generic processes. Social order is simply group solidarity writ large. The idea that state control is isomorphic to lower-level group control is not a new theoretical insight (Eckstein [1966], 1992), but it has never before been incorporated in theories of state formation and maintenance.

Doing so raises three related questions. Why are pre-existing social groups necessary for the formation of state? Why do such groups remain extant even after the state has formed? And what prevents unattached individuals from forming and maintaining a state? The answers hinge on the cost of maintaining a state, which is minimized when groups intervene between individuals and the state. Let the efficiency of maintaining social order be defined as the ratio between $\gamma_s$ (where $S$ refers to the society as a whole) and the number of individuals in the society. Then the more resources that must expended to maintain order, ceteris paribus, the less efficient the society’s control institutions are. It then can be shown that efficiency increases with the size of constituent social groups, even if the total population of the society remains constant and the costs borne by the groups increase proportionately with the number of individuals they encompass.

Let $N$ be the total population of the society in individuals, and $n$ be the number of groups into which the society is divided. If no groups mediate between individuals and the state, then $n$ will simply be equal to $N$. Otherwise, the average size of groups will be $N/n$. If each individual in the society is expected to contribute some uniform amount $c$ through the auspices of her group, and all groups in the society are of equal size, then the total cost borne by the average group will simply be $c(N/n)$. Let $\gamma_s$ refer to the (infinite-level) aggregated costs of order in the entire society $S$. Then $\gamma_s = \Phi(n(cN/m) + \gamma_{s}^m/p_s) = \Phi((cN + \gamma_s)/p_s)$. Since $\Phi$ is a monotonically increasing function, $\partial \gamma_s/\partial n$ has the same sign as $\partial((cN + \gamma_s)/p_s)/\partial n$. This is equal to

$$\frac{\partial \gamma_s}{\partial n} \frac{1}{p_s} + \frac{\partial \gamma_s}{\partial n} \frac{1}{p_s} (cN + \gamma_s).$$

The second part of this sum is positive, since $p_s$ is a positive and monotonically decreasing function of $n$ while $c$, $N$ and $\gamma_s$ are all defined to be positive. Since $1/p_s$ is by definition positive, the first part of the sum also has the same sign as $\partial((cN + \gamma_s)/p_s)/\partial n$.This calculation allows $\partial \gamma_s/\partial n$ to be decomposed to the sum of an infinite series of terms, each of which is positive. Thus the total is also positive, and $\partial \gamma_s/\partial n > 0$.

Hence, given equal individual-level obligations, the presence of groups mediating between individuals and the state reduces the state’s costs of maintaining order (Hechter, Friedman and Kanazawa, 1992; Hechter and Kanazawa, 1993; Hechter, 1993). Group solidarity contributes to social order by regulating members’ behavior, thereby constraining their opportunities to engage in anti-social behavior. Further, states in societies with fewer and larger groups at the sub-societal level will be more efficient at maintaining order than those in societies containing several groups, because they require a smaller outlay of resources per individual for monitoring and sanctioning.

**State-Level Versus Group-Level Costs of Order**

A critic of the previous analysis might argue that it biases the results by only considering the state’s costs of maintaining order. The state’s savings from group mediation, this critic would note, are simply being passed down to the groups, which have to take responsibility for maintaining solidarity among their members. Hence, the cost to society as a whole may be affected differently by this group mediation than the cost to the state.

In response, we argue that the state’s costs are particularly important because they are the final determinant of the existence of social order. In addition, highly solidary groups may be capable of demanding contributions from their members that the state cannot, particularly when it is a new entity on which individuals are not highly dependent (see Stinchcombe, 1965, p. 148 on the liability of newness). And due to their scale, groups are much more capable of monitoring and sanctioning than the state can ever be (Hechter, Friedman and Kanazawa, 1992; Hechter and Kanazawa, 1993). Nonetheless, the total social costs at all
levels are important since they determine the overall burden borne by individuals. Rational egotists always seek to minimize these costs. Even when total social costs are taken into consideration, however, it can be shown that group mediation reduces the costs of maintaining order.

An appropriate measure of the total resources necessary to maintain order in society $S$ is the total resources needed to maintain order between the various groups in society plus the total of the resources needed to maintain solidarity within each group. Let $\Gamma_S$ be this quantity, which is defined recursively as $\Gamma_S = \gamma_S + \sum \Gamma_s$ for all members $s \in S$, and where $\Gamma_s = 0$ when $s$ is a solitary individual. As the number of groups decreases in a society of fixed population, the average size of groups increases, and the cost of maintaining group solidarity rises as the cost of maintaining order between them falls. By this logic, although incorporation into a few large groups lowers the amount of resources devoted to maintaining order at the societal level, it simultaneously raises the amount of resources devoted to maintaining solidarity at the group level.

Which scale of nested control institutions is most conducive to the maintenance of social order at the lowest cost? The issue is complex, since no arrangement is most efficient at all levels of analysis. Consider the simplest case, where hierarchies are structured such that all groups at each level contain some uniform number of members. This number, $n$, can be seen as some measure of the shallowness of the chains of control in a society. Where $n = N$, a mass society having only a single level of authority between the state and atomized individuals obtains. At the other extreme, the deepest hierarchy will be one in which no group has more than a few subgroup members.

As noted earlier, the value of $\gamma_S$ at the state level is $\Phi((cN + \gamma_S)/p_S)$. At the next layer of the hierarchy, the cost of maintaining order among subgroups within each group $s$ is $\Phi((cN/n + \gamma_S)/p_S)$. Hence the total cost of maintaining order at this level will be $n\Phi((cN/n + \gamma_S)/p_S)$. A similar logic can be used to calculate costs at lower levels of control. In general, $\Gamma$ will be the sum of the costs at all of these levels.

While it is difficult to characterize the value of $n$ that will minimize $\Gamma$, it is possible to show that there exist plausible $\Phi$ and $p_S$ so that group mediation between the state and individuals can reduce $\Gamma$. For instance, suppose $\Phi$ is a linear function. Then $\gamma_S = \Phi((cN + \gamma_S)/p_S) = V(cN + \gamma_S)$, where $V$ is simply the slope of $\Phi$, divided by $p_S$. Then $\gamma_S - V\gamma_S = VcN$, and $\gamma_S = (V(1-V))cN$. At the next layer of hierarchy, the cost of maintaining order among each subgroup will be $\gamma_S = V(cN/n + \gamma_S) = (V(1-V))cN/n$. However, since there are $n$ groups at this level, the total cost of maintaining order at the level will be $V(1-V)cN$, as well. A similar logic holds at lower levels; hence it can be concluded that such uniformly nested hierarchies will have equal total costs of maintaining order at each level all the way down the hierarchy, and that this cost will be equal to $\gamma_S$.

Thus the total (aggregate first- and higher-order) cost of maintaining order is simply equal to the number of layers multiplied by the cost at each layer, $\gamma_S$. Since each layer reduces the size of groups by a factor of $n$, the number of total layers will be such that $n^l = N$, where $l$ is the number of layers. Therefore (ignoring indivisibilities) the total cost of maintaining social order will be minimized when $\log(N)/\Phi(cN + \gamma_S)/p_S$ is minimized. As a function of $n$, this is minimized when $(\ln(N)/\ln(n))\Phi((cN + \gamma_S)/p_S)$ is minimized, which in turn is minimized when $(cN + \gamma_S)/p_S \ln(n))$ is minimized. Using a similar method as above, this can be seen as the sum of an infinite series of terms, each of which is minimized when $1/(p_S \ln(n))$ is minimized, that is, when $p_S \ln(n)$ is maximized. The first-order condition is $\partial(p_S \ln(n))/\partial n = 0$. This is true when $\ln(n) = -\frac{1}{e} \Phi'(p_S \ln(n)) = (1/p_S)/\Phi'(1/p_S)$. More concretely, it is clear that $n=1$ is not optimal, since $\ln(1)=0$. Hence, up to some point mediation between individuals in the state will reduce the total societal costs of maintaining order. However, for sufficiently steep $\Phi$ functions $\text{vis-a-\textit{vis}} n$, it is clear that $n=N$ will not be optimal either, and there will be an interior solution.

The steepness of the relationship between group size and the cost of monitoring and sanctioning has a tendency to reduce the optimal number of members in groups at each layer. This implies that when groups grow larger and monitoring becomes more difficult, social order can be maintained most efficiently overall through narrow and deep chains of authority. When monitoring is less problematic, broad and shallow chains of authority are less costly.

Narrow, deep chains of authority seem to characterize the structure of small supergroups that are particularly difficult to monitor. This is one reason why underground political and military organizations tend to organize themselves into small cell structures (Chai, 1993). A similar structure characterizes societies lacking advanced communications, where distance and cultural diversity impede monitoring outside group boundaries. This analysis may explain how some societies are organized into a few groups just below the state level having narrow deep chains of control (as in feudalism), while others are divided into more numerous groups with broader and more shallow lines of control (as in modern unitary bureaucratic states). It may hence offer an explanation
for the transition from patrimonial to bureaucratic state forms as described by Weber ([1922] 1968) and Tilly (1990), among many others. Finally, this analysis has a clear, if indirect, relationship to organizational theories on the span of control and the configuration of hierarchies. For instance, it formalizes the idea that as monitoring becomes more difficult (as a function of size, or due to the complexity of tasks at hand), lines of authority tend to become longer and narrower (Blau, 1968).

The upshot of this analysis is that *some* sort of group mediation – or *indirect rule* – is always necessary to minimize the cost of order in large societies. Yet mediating groups that are too large given the constraints of the environment may actually have a negative effect on order, primarily because of the large amounts of resources they must expend to maintain internal solidarity.

**Dependence, Coercion and Social Order**

Given that a state is simply a group composed of members that are themselves groups, the arguments developed above about the effects of dependence and the possibilities for coercion in groups also apply here. At the time of state formation, the level of dependence of groups on the state will be equal to the difference between amount of net benefit they gain from the given state and that to be gained from a rival state, or under anarchy. This level of dependence puts an upper ceiling on the amount of sanctions that the state can impose on noncompliant groups, which in turn puts a ceiling on the extensiveness of obligations that can be demanded from these groups. As is the case in a group, higher levels of dependence can actually allow states to produce more valuable joint goods for member groups. High levels of dependence among one set of members also may promote coercion, however, as such groups may be able to form a coalition and mutually enforce each other’s contributions towards coercing nondependent groups to remain in the state.

Besides these society-level manifestations of simple group phenomena, however, further issues are raised by the complex, nested structure of the state. Dependence is important to the maintenance of social order and possibility of state coercion. Social order may not be attainable unless the members of intermediate groups are superdependent (in the sense used above). Without superdependence, it is impossible to use group control to enforce individual compliance with additional, societal obligations. Hence, social order should vary with member dependence of mediating groups. In turn, dependence is fostered by a long history of interactions. Japan, the most ordered large industrial society, appears to be a case in point (Hechter and Kanazawa, 1993).

Likewise, coercion at the societal level depends on superdependence in coercing groups to ensure their members’ contribution to the coercive apparatus. Coercion also can be fostered by insufficient dependence of the members of coerced or oppositional (Hechter, 1993, pp. 8–10) groups, since the inability of such groups to participate in a planned simultaneous exit (such as secession) can reduce the amount of coercion required to keep such groups under control. The major implication is that states in which there is high variance in member dependence across groups will be most prone to coercive rule. This analysis points out the need to distinguish between social order in general and voluntary order in particular. Dependence always fosters order, but the uneven distribution of very high levels of dependence can lead to coercive forms of order.

**CONCLUSION**

We have presented a formal model to investigate some aspects of group solidarity and its relationship to the attainment of social order. In doing so, we have also provided a parsimonious explanation of the formation of the state. Along the way, a number of hypotheses have been generated:

- High levels of group obligation interact with group size to raise the costs of maintaining group solidarity. Hence groups that demand extensive obligations of their members will have fewer members than groups with less extensive obligations. The societal analog of this proposition is that states that demand extensive obligations – highly centralized ones – will tend to have deeper control institutions than those which demand less extensive obligations – federal ones. Because states that demand extensive obligations must employ more monitoring, they have to rely on smaller intermediate groups.
- Dependence sets an upper bound on the extensiveness of obligations any group can demand of its members. Further, dependence is a product of a long history of interactions. Since more extensive obligations often lead to more valuable joint goods, groups having long-standing, stable boundaries, and whose members have fewer exit alternatives, should produce the most valued joint goods.
Because a group's ability to contribute to social order is also a function of the dependence of its members, states are most likely to emerge on the shoulders of groups with long-standing boundaries and geographical and social impediments, such as ethnic groups.

- Whereas dependence is conducive to order, it is also conducive to coercion both within and between groups. Groups with highly dependent members are able to mobilize these members to exploit less dependent members by forcing them to stay in the group. Superdependence also enables some groups to coerce others to remain within a state control system. Thus coercive order should be most common in societies where part of the population belongs to long-standing communal groups and another part does not.

- The presence of solitary groups mediating between the state and individuals usually reduces the state's costs of maintaining order. Further, the larger the size of these mediating groups, the more the burden of maintaining order falls on groups rather than the state. This advantage is particularly important where individuals and groups have low dependence on the state and thus high autonomy from it; as a result, the state is incapable of imposing extensive obligations. Hence newer states with less autonomy (in general, those with fewer resources) will be particularly likely to be based on the mediation of larger groups and to rely on deeper lines of control.

- Finally, the total costs of social order are related in a complex way to the structure of nested control in a society, though some form of mediation will always be necessary to minimize costs. Apart from group dependence, the crucial factor determining the optimal structure for maintaining social order is the relationship between group size and the cost of monitoring group behavior. The steeper this relationship is, the deeper the optimal lines of control. Because this steepness is greatly influenced by the infrastructure of communication and surveillance, one should expect that socioeconomic modernization will tend to lead to shallower lines of control over time.

Thus, a model of group solidarity based on individual actors can not only account for the development of order among groups, but it can also explain the emergence of the state. This model generates new solutions to long-standing debates.

Much more work remains to be done, however. Most important, we have said nothing about how the benefits generated in a group or supergroup are distributed among its members. This modeling task lies ahead. We conclude by broaching few more substantive topics that will require analysis in the future: the development of state autonomy, the effect of the international system and the possibility of parallel membership in multiple groups.

**State Autonomy**

The state may originate as a set of agreements between social groups, but eventually it becomes an actor of its own that receives resources from groups, provides a wide variety of public goods and enforces laws (Skocpol, 1985; Krasner, 1984). The state not only interacts with groups; it can alter the nature of these groups. Whereas there is growing interest in the motives of the state as an actor and its effects on society (Levi, 1988; Bates, 1981; Magee et al., 1989), this literature generally takes the state as given rather than accounting for its origins.

The present model could incorporate state autonomy by positing that the formation of a state implies the creation of a special group that assumes responsibility for monitoring and sanctioning for the society. Initially this state-actor simply ensures agreements between a set of pre-existing groups. Later, however, the state-actor may attempt to modify the chain of authority, and to restructure agreements for the state-actor's benefit. Generally, state control is supported by creating a group that has special access to resources and coercive power. Thereafter, the remaining groups will attempt to provide sufficient resources to make the state effective, but not enough to allow it to coercively control them.

**The International System**

The present model provides a hermetically sealed view of social order. In common with most traditional models of social order, it considers a society to be a closed system insulated from exogenous forces. Despite this, the model can readily be extended to encompass interaction in the international system. Interaction between states in the international system quite properly can be considered isomorphic to the interaction of groups in a given territory. An analogous relationship exists between the internal order of societies and the ability of states to negotiate a common set of rules for interaction. In such a model, the difficulties of forming an all-encompassing world superstate can be linked to the initial inability of such a state to engender high
levels of dependence among its member states, as well as the large number of existing states in the world, which makes the maintenance of order even at the top level relatively costly.29

**Multiple Group Affiliation**

Perhaps the most dramatic simplification in the present model is the assumption that social control institutions have a tree-like shape, with each individual and group located on a single line of control ultimately leading to the state. This is far from an accurate portrayal of most societies, particularly modern ones, where individuals simultaneously belong to several groups which are not nested in concentric rings or are only partially nested (Simmel [1922], 1955). The idea of a tree-like control institution was adopted as a simplifying assumption, one that presumably captures the fundamental characteristics of even more complex societies. Bringing more complex patterns of control into the model is quite possible, however, albeit with some loss in parsimony. In an altered model, compliance by individuals with obligations at the state level could take place through several parallel lines of mediation. In such a model it would be possible for compliance to occur along one line and not along another, thus allowing for partial compliance as well as compliance and noncompliance. This modification would make the calculation of optimal patterns of intermediation inordinately complex, but it would not change the basic calculus of obligations, monitoring and sanctioning that drive the results found here.

**Group Member Heterogeneity**

Much of the analysis above assumes for simplicity that all members of a group are identical in terms of the contributions they make and the benefits they receive from the group, the net utility that they gain from these, and the opportunities they have outside the group. However, the model is quite compatible with extreme heterogeneity among group members, and indeed the assumption of heterogeneity is implicit in the discussion above of differential dependency and its relationship to coercion. Depending on the circumstances, group heterogeneity can have either a facilitating and inhibiting effect on member cooperation (Marwell and Oliver, 1993). Among the main points of interest for this model are the ways in which group heterogeneity affects variations in members’ required contributions. While the model above treats required contributions as exogenously determined, they are clearly a product of bargaining that goes on among group members.30 In particular, dependence can be seen to have a dual effect on the nature of that bargaining over relative contributions. Both nondependent and dependent members will have sources of leverage: nondependent members may credibly threaten to exit the group, yet dependent members may credibly threaten coercion in response. In order to determine how all this affects the nature of the control institution there is a need for a dynamic model of the bargaining process.

Although the present model offers far from the last word on these complex issues, it suggests that a host of significant questions can be explored through further formal analysis of the relationship between solitary groups, the state, and social order.

**ACKNOWLEDGEMENTS**

Earlier drafts of this essay were presented at the Conference on the Role of Institutions in the Public Sector, 9–10, November 1995, Netherlands Interuniversity Institute of Government, Oosterbeck, The Netherlands at the 1995 Annual Meetings of the American Sociological Association in Washington, DC, and in the Department of Sociology of the University of Stockholm. We are grateful to Göran Ahrne, Yoram Barzel, Patrick Doreian, Thomas Fararo, Satoshi Kanazawa, Edgar Kiser and Karl-Dieter Opp for their comments.

**NOTES**

1. This large body of literature focuses on the role of formal institutions in structuring interactions among rational actors. See, for instance, the review essays by Moe (1984) and Shepsle (1989).
2. For a rational-choice application premised along these lines, see Hechter and Kanazawa (1993).
3. There have been some attempts to generate rational-choice models of such plural society arguments; see Rabushka and Shepsle (1972). More recent versions have stressed the problem of "commitment" among such groups experience in the absence of an effective state. See Fearon (1994) and Weingast (1994).
4. To this point in the analysis, we employ rational egoistic behavioral assumptions. Later, we consider the effects of altruistic preferences, but these are not fundamental for our results.
5. These contributions require groups to have more than one member, since a single member would be able to produce purely private goods and thus would face no free rider problem. A "group" of a single member is a trivial case, and such groups can be treated identically to the individual who comprises them.
6. Of course, it is implicit in the analysis that there be benefits $b_k$ exceeding $c_k$ in order for group membership to be worthwhile in the first place. We will discuss even more stringent conditions for voluntary membership in the section below on dependence.
mechanism - see discussion at end of paper for some suggestions on how such analysis might proceed. Thanks to Satoshi Kanazawa for bringing up this issue.

21. As implied by the term hierarchy, the nesting of groups follows a simple tree structure, where each group is a member of at most one group at the next highest level, and where group memberships at the same level do not overlap. Individuals will be at the bottom of this hierarchy, and the simple noncyclical nature of the hierarchy will imply that each individual will belong to only a single chain of group membership.

22. Incidentally, in this theory it is not necessary that this atomic level occur at the same level of the hierarchy for each chain of authority. Thus it is possible for some members at the same level of the structure to be individuals and others to be groups consisting of more than one individual.

23. In some historical circumstances, states created groups de novo in order to extend their control. Thus the Spanish Crown established the Mesta, a shepherd guild, to facilitate its revenue extraction (Klein, 1919).

24. What about predatory (organized criminal) and oppositional (potentially revolutionary) groups, whose very existence obviously represents a potential threat to social order? Recall that the possibility of predation is what motivates the establishment of the state in the first place. Predatory groups are best viewed as the consequences of incomplete social order, but due to monitoring costs in large societies it is unlikely that they can ever be eradicated (for an excellent analysis, see Barzel, 1995). Oppositional groups have a negative dependence on a society, and thus have an interest in overthrowing the prevailing order. Even then, however, their existence does not increase the cost of maintaining order in our theory, since absent these groups acting as mediating entities the state would have to impose compliance on a set of disaffected individuals directly, which would entail even greater monitoring costs. The cost of sanctioning in this model is related to ensuring positive contributions to the state, not preventing negative contributions, such as military rebellion. Note, however, that organization of individuals for the former does not necessarily entail organization for the latter, or vice-versa.

25. If this were not true, it would be possible to eliminate any cost of order in a society by creating a single all-encompassing group immediately below the state level. As long as within-group order must be maintained, however, such a step will have no effect on the cost of order, since it will effectively transfer all the costs from the state level to the group level without any other change in configuration.

26. Since the number of members at each level is the same, we can from here on refer to the probability of detection at all levels as \( p_k \).

27. Tilly (1990) has an excellent discussion of the importance of indirect rule in the early development of the modern state, but he considers it irrelevant in the contemporary situation. This view is quite mistaken, however. For reasons that this theory elucidates, indirect rule continues to be essential to the production of social order even in the more developed states today.

28. The level of group obligation may be endogenous, however. Wherever it is easier to achieve more of it will be demanded, and the number of members will increase.

29. This may also explain why a super-state unit such as the European Community was formed as out of a smaller collection of mutually interdependent states.

30. Thanks to Yoram Barzel for bringing up this point.

REFERENCES


Fears, James (1994) Ethnic warfare as a commitment problem. Unpublished manuscript, Department of Political Science, University of Chicago.


Hechter, Michael (1993) From group solidarity to social order. Unpublished manuscript, Department of Sociology, University of Arizona.


THE STATE AND SOCIAL ORDER


Skocpol, Theda (1985) Bringing the state back in: strategies of analysis in current research. In Peter B. Evans, Dietrich Rueschemeyer and Skocpol (Eds.), Bringing the State Back In. Cambridge: Cambridge University Press, pp. 3–44.