

November 19, 2004

Float Together/ Sink Together?
(The Effect of Connectivity on Power Systems)

by
Richard E. Schuler
Prof. of Economics
Prof. of Civil & Environmental Engineering
Cornell University
422 Hollister Hall
Ithaca, NY 14853-3501
<res1@cornell.edu>

The recent mantra for reorganizing power systems in the U.S. has been to extend the geographic scope of control centers to span several states, utilities and/or grid operators, initially for the purpose of expanding the range of economic transfers and more recently to improve operational reliability, in both cases through the reduction of “seams” at the borders of control areas. In the early days of electric deregulation this push for coordination was in the guise of forming four to five Regional Transmission Organizations (RTO), combining existing power pools and Independent System Operators (ISO), that might dispatch power at least-cost over wide regions of the country. The Federal Energy Regulatory Commission (FERC) also proposed a standard market design (SMD) for all control areas so that neighboring entities could exchange power more effectively, but this initiative has fallen victim to massive states’ rights battles (Whatever happened to the Commerce Clause of the U.S. Constitution?). Following the August 14, 2003 Northeast blackout, similar calls for far greater regional coordination have been based upon the perceived benefits in terms of greater reliability and reduced susceptibility to cascading disturbances across control area borders.

Currently, the power system(s) in the U. S. is a hodge-podge - - institutionally, economically, physically and in terms of regulatory oversight. It is the epitome of nation-wide de-centralized decision-making about a set of systems that are, nevertheless, highly centralized locally. This analysis reviews these seeming inconsistencies and examines the likely consequences for reliability. Conceptually, it compares strongly coordinated network systems vs. decentralized loosely coupled systems as applied to the vulnerability of power grids to catastrophic collapse. As an example, would the Northeast Blackout of Aug. 14, 2003 have been limited or more widespread (and therefore tougher to restore) had the PJM and New England ISOs not separated from New York prior to New York’s protection system isolating it from Ontario? How does this experience affect the way we design future systems to improve their sustainability in the face of both natural and terrorist threats? How might that prospect affect the terrorists' targeting?

From the perspective of critical infrastructure, none seems as essential for the working of modern societies as the electricity grid. While the telecom, transportation, water and sewer and financial networks are highly interdependent with the electrical network, on a first principles basis the others cannot be sustained for a long period without reliable electricity supplies. As an example, following the August 14 blackout, several cities lacked water supply until the electric service was restored. And in New York City following the September 11 attack on the World Trade Center, water was available to fight the fires only because diesel-powered fireboats pumping water from adjacent rivers were pressed into service. Telecom service in NYC was also rapidly restored in large part because of diesel-powered emergency generators, but with prolonged outages of the electricity grid, how's the fuel to get in without the signals, powered by electricity, to unsnarl traffic? With such an ill-coordinated electricity system across geographic and institutional boundaries, it would appear at first glance that this is by far our most critically vulnerable system. While that is true for local pockets of customers, the good news is that on a wide geographic scale, hodge-podge and loose coupling may be beneficial! Short of all-out, ongoing warfare, it is difficult for even a coordinated terrorist action to bring large contiguous sections of the power grid down for more than a day or two. Hurricanes seem to do a much better job than can malevolent people.

After reviewing the design and operating practices for electricity systems throughout the U.S., simple agent-based simulations are used to illustrate essential principles about the relationship between the size of an organization (e.g. the number of local activities or control areas that agree to coordinate their actions and/or are spanned by an ISO/RTO) and the number of external connections (e.g. interconnections with other entities whose objectives are autonomous, like other ISOs or threatening activities) where individual local control areas "learn" how best to respond to system insults. Overall, the greater the number of non-cooperating external pathways, the larger the organization should become in order to enhance reliability (performance). However, the outcome hinges in part upon the expected duration of the threatening environment. For short time horizons, it is useful to build large interconnected entities so a larger number of experiences can be shared, but as the planning horizons is extended, the optimally-sized organization grows smaller, even in the face of many potential external insults, as the improvement of performance of individual agents outweighs the "confusion" created by too many tightly-linked partners.

1. Primer on Electricity Systems

Most major electricity systems use alternating current (AC) because that is essential for transforming voltages (the energy potential) from one level to another, and high voltages, in turn, are required for the economical hauling of electricity over long distances (with low losses). But AC systems cannot modulate the flow of electric energy over particular paths, unlike water and natural gas flows that can be adjusted by turning valves; instead, the flows between generators and users and the paths selected are governed by the laws of physics (Mr. Kirchoff's). Furthermore, since economical generation methods are usually very large in scale, the suppliers are concentrated at particular points (frequently

near fuel supplies or where environmental impacts are minimal, and always near water), and multiple transmission lines are erected to provide alternative paths that enhance the reliability of connecting those low-cost generators with concentrations of customers in urban areas.

An illustrative schematic diagram of a bulk power system is shown in Figure 1. Here transformers step the voltage of power produced by generators up to higher levels in preparation for long distance hauling over transmission lines (the solid lines). Similarly, step-down transformers reduce the voltage in preparation for delivery to the local distribution system. Frequently these transformers are located together with switches and connecting facilities called busses in substations (encircled with dashed lines in Figure 1). These transformers are large and are located on the ground or underneath in vaults, but the connecting busses are usually located overhead on a steel superstructure. There are also many switches located in each substation that can connect or disconnect the various transformers and lines at each of the junction points, and some of them operate automatically when a problem is “sensed” on the system, much like the circuit breakers in the switch box in every home. However, since this is an AC system, those switches are either entirely open or closed, and the apportionment of power flow among those lines when there are parallel paths is governed entirely by Kirchoff’s Laws.

(Insert Figure 1)

This simple system is constructed so that it satisfies the single contingency design criterion for bulk power system in the U.S., that is, the system can withstand the interruption of any single transmission line or generator and still be capable of delivering power. In fact, in the hypothetical system illustrated in Figure 1, every load center can continue to be served despite the loss of any single generator (if those generators are each sized with sufficient excess capacity, all loads might also be met with the loss of any two generators), and the same is true with the loss of any one of the three major transmission lines (those connecting 1-2, 2-3 or 2-3).

To maintain this reliability, the location of switches and breakers is important. As an example, if line 2-3 is interrupted, Load C would be interrupted, unless there is a switch on either side of junction point “0”. If so, then a break in line 2-3 between 0 and 3 could be isolated, and Load C would continue to be served through the 3-0 portion of line 2-3. As drawn, however, all of the loads except Load-D would be lost if their major connecting substation were incapacitated. As examples the loss of S-6 knocks out Load A, the loss of S-1 disconnects Loads A and B, and elimination of S-7 denies power to Load C. Load D, however, must have two substations eliminated (particular combinations of S-1,3,4 & 5) in order to be denied service, and so its designed level of reliability is significantly greater than the other loads. In fact, major cities and/or large industrial customers are situated like Load-D; however, depending upon their internal electrical connections at the distribution voltage level, portions of their loads may be disrupted by the loss of a major transmission substation.

In fact the system design and reliability criteria vary appreciably at the distribution voltage level from that at the bulk power system level. With the exception of large, high-density urban areas like NYC, the lower voltage distribution system (the load side of substations in Figure 1) is routinely laid out with radial feeds to customers from the substation. These are the lines on the cross arms of poles that line highways and city streets or are underground adjacent to them. Because of the radial configuration, when a distribution line is interrupted, all customers along it are without service. However, if these lines are skillfully configured spatially so that the ends of two lines are in close proximity, some customers may be reconnected to the system before the original cause of the line outage is repaired, simply by opening and closing switches along these lines, and back-feeding those customers beyond the short-circuit on the interrupted line.

Nevertheless, because the distribution system is usually exposed to many more insults as a result of its ubiquitous presence along nearly every street and highway (the exception is in major cities where the distribution system is configured differently as an underground network), and because of its radial configuration without instantaneous alternative paths of supply, approximately 80 percent of all power disruptions in the U.S. are caused by distribution system failures. The remaining 20 percent are attributable to failures in the bulk power system illustrated in Figure 1. However, since all major regional blackouts are failures of the bulk power system (far more than a single contingency is experienced simultaneously), studies of catastrophic failures focus on the bulk power system. Only in the case of widespread natural catastrophes like hurricanes or ice storms, are the sources of disruption usually at both the distribution and bulk power levels. In these cases, however, the bulk power system is usually restored to full service in a day or two, primarily because of its network configuration with parallel paths, and the prolonged customer outages of more than a week are usually the result of the multiplicity of distribution failures that must be repaired one by one until all customers can be restored to service.

The exception is in large cities like New York where because of the spatial density of both demand and of the distribution pathways along the local streets, not only are all distribution facilities placed underground where they are less exposed physically, they are configured in a grid as a tightly coupled set of networks with many parallel paths and multiple sources of bulk power supply. Under this distribution configuration, the loss of any one or two lines results in hardly a blip in the service to any customer, and this reliability is accomplished automatically without dispatching operators to open and close switches, as is the case with radial distribution systems. Obviously, the cost of providing electric service is also much greater when this underground network configuration is used, but the extra cost is proportionately lower as the spatial concentration of demand increases, as in megalopolises like New York.

2. Regulatory and Institutional Structure

So who's in charge of this complex system? The answer is: lots of entities and institutions, and therefore in effect, no one! Although the various private and public facility owners and operating entities (privately owned utilities, municipal electric

companies, state and federal power agencies and rural coops), the coordinating power pools, Independent System Operators (ISO) and Regional Transmission Organizations (RTO), and the state PUCs and PSCs (regulatory bodies), as well as the Federal Energy Regulatory Commission (FERC) all have a role to play, none of these entities has clear overall authority to provide (or order to be provided) the necessary facilities, connections and coordination that would enhance overall system reliability or lead to regional economy of service! That is why after the 1965 Northeast blackout the National Electric Reliability Council (NERC) was formed, together with its regional councils, but the operating and design guidelines and recommendations of NERC are voluntary because it is a voluntary industry-formed organization. And the major push for the deregulation of exchanges of bulk power supplies that occurred in the 1990's was motivated by attempts to achieve more economical regional costs of bulk power supplies through market incentives, but many regions of the country are still balking at the introduction of systematic (and uniform) exchange mechanisms.

So while FERC can nudge the creation of ISOs and RTOs (both organizations are agents of FERC) to create larger market areas and to operate the transmission systems that undergird these market areas reliably, so far they have been unable to withstand the political backlash of ordering their formation. Particularly vehement has been the local differential reaction to imposing a uniform structure to these markets (FERC has outlined a Standard Market Design (SMD) in very broad, flexible terms, but even that has drawn severe derision from some regions of the U.S.) so that exchanges can be made efficiently and reliably across the borders of the control areas. Furthermore, FERC has recently been rebuffed in its attempts to order (provide incentives for) the construction of transmission links that might actually allow the power to flow physically across these borders.

Similarly, in the area of establishing and maintaining reliability standards, NERC is toothless and therefore unable to mandate and enforce compliance with its issued guidelines. And so despite an heroic analysis of the causes and faults of the August 14, 2003 blackout in which NERC identified many instances of non-compliance with its rules, it has no authority to impose sanctions. This leaves FERC searching for threads whereby it might impose sanctions on hardware owners and operators for non-compliance, and many knowledgeable and concerned professionals around the nation calling for Federal legislation that would make compliance with NERC standards and guidelines mandatory. In fact, most state public service commissions do have the authority to impose mandatory performance guidelines on the utilities that they regulate and to back them up with penalty actions in subsequent rate proceeding if there is inadequate compliance. In most instances, they can also authorize the construction of needed new facilities. The problem with each state's authority, however (which may not be uniform across states), is that it cannot reach beyond its limited political borders, and as the August 4, 2004 blackout demonstrated, many of these events are multi-state requiring regional solutions.

3. Operation, Control and Reliability Philosophy

Operational control of these complex systems is currently in the hands of a system operator who together with her staff oversees the operation and dispatching of power within her control area (power pool, ISO or RTO). In addition, each of these regional control centers will call upon operators at smaller area control centers that are staffed by individual utilities and/or at individual generating stations to carry out orders to open or close switches and to increase or decrease the supply from any particular generating unit. Usually, this dispatch is designed to minimize the total cost of supplying all power demanded by customers, subject to the available generation capacity, flow limits on individual transmission lines, and maintaining adequate service quality (frequency at 60 cycles and design voltage levels); otherwise the performance of users' and generators' equipment, indeed its survival, might be compromised. This constrained optimization is so complex that it must be solved by a computerized routine, usually every fifteen, but increasingly every five minutes. The costs of generation for each unit and their available capacities are furnished by the suppliers in a regulated utility power pool framework or by price-quantity offer schedules from potential suppliers in a market context. In either case, excess generation (above the power demanded) that is equal to the size of the single largest generator then operating is always kept running so that the system's load can be matched should that largest unit fail, with a lag of no longer than five minutes (operating reserves).

In planning and ordering these dispatches, the system operator must know what units are available to be called upon in an emergency and which lines are out of service, so that if any contingency occurs on the bulk power system, she can immediately have a revised optimal dispatch computed. Because of Kirchhoff's Laws, an operator cannot dictate over which line an ordered increase in generation will flow. As an example, if there is increased load at L-B in Figure 1, and it is scheduled to be served by the next highest cost producer at G-2 over transmission line 2-1, in fact the increased generation at G-2 will most likely flow over both parallel paths, 2-1 and 2-3-1. If the capacity limits are being reached on line 3-1, the amount of generation reaching L-B from G-2 may be limited, even if line 2-1 has spare capacity. Kirchhoff's law of equalizing voltage drops across parallel paths will prevail. In this example, the only way the operator can guarantee that the power might flow over line 2-1 is to open the switches at both ends of line 3-2; thereby taking that line out of service. The flow on any individual line in this network of parallel paths cannot be modulated without the addition of expensive new technology. And, having parallel (redundant) paths is essential for maintaining the reliability of the bulk power system.

This example also illustrates how crucial it is for the system operator to have accurate up-to-date information about the operation and condition of all equipment, including which lines are or will be out of service. Where control areas are interconnected (as they are across ISOs and RTOs in the northeast), the informational needs are more extensive, and it is equally important to know what's happening in neighboring control areas as well. That was not the case on August 14, 2003.

In addition, many things that happen within these power systems is too fast for human intervention (it may be too fast for centralized computerized analysis and response as

well). Thus an automated back-up protection system is in place, triggered by sensing devices at particular locations in the grid that send signals through relays causing circuit-breakers (automated switches) to open and close. Although these “trip-signals” are planned and simulated on a system-wide basis ahead of actual events, once the devices are set, they operate autonomously on a totally decentralized basis. As an example, were a transmission tower on line 1-3 in Figure 1 to collapse, the resulting short circuit would cause sensing devices to “see” an in-rush of electric current from both points 1 and 3, and the associated relays most probably would be set to open the breakers on the busses at points 1 and 3 that feed line 1-3, thereby isolating the problem. In this case, because of the redundant configuration of the bulk power system all loads would continue to have service following this contingency.

If instead, a hot air balloon were to land by accident on a transmission bus at point 0 providing a path for electricity to flow to ground, sensors at points 2 and 3 might signal associated relays to trigger the breakers to open at points 2 and 3 on line 2-3, once again isolating the problem. But in this case load L-C would be disconnected from the system, since no redundant path is available to serve it. However, because it is probable that a hot air balloon falling on the bus at point 0 might eventually slide off of the facility (or burn up because of the heat generated by the electricity flowing through it to ground), the short circuit might last for only a few seconds, or even a fraction of a second. In that case, it would be unfortunate for customers at L-C to be out of service for an extended period of time, waiting until a crew could be dispatched to inspect the source of the problem at point 0, and if cleared without any further structural damage, to reclose the breaker. That is why circuit breakers are installed in many locations with automatic reclosing features. Frequently, the breakers are set to test the line twice after opening automatically the first time. With a short delay, the breaker closes automatically after a preset interval (one or two seconds, as an example) and then opens again if the short-circuit is still detected. In some cases the breaker is programmed to try to close a second or third time after successively longer waits, but usually on the third try, it is locked open waiting human intervention. In this way, failures that might be transitory, like lightning strikes or tree limbs blown against a line, are interrupted, but then restored automatically and quickly if the original insult has moved on without causing permanent damage.

Human judgment and decision-making determines the nature of detection devices and relays that are installed and where and under which measurable conditions the breaker will open (e.g. massive change in current flow, impedance on the line, voltage, etc., since the sensors can’t actually “see” a short circuit). Human design also determines how rapidly the breaker should open after detecting an abnormal condition. A fast response reduces the chances of damage to facilities or people along the line, but it also increases the chances that customers will be interrupted if the phenomenon measured was only transitory, or worse, a false signal. As long as the breaker remains closed following a persistent assault on the line, the system will continue to feed tremendous amounts of energy into the short-circuit with a consequent increased probability of destroying not only the object that interfered with the line, but also the line itself and its supporting structures. Erring on the side of a very rapid response minimizes the chance of damage to

facilities, but it increases the probability that many customers will be inconvenienced at very high costs to them, even when the source of the breaker operation was transitory.

The trade-off is between trying to maintain the integrity of the entire system by keeping the breaker closed or keeping the potential physical damage to the system to a minimum by interrupting the power flow rapidly. This is a precise example of a trade-off in protection philosophy: how long do we keep all of the lifeboats lashed together, even though several have holes in their hulls, so that everyone might survive, versus when do we cut those damaged boats loose so we all don't sink together?

There are further ramifications to this choice of breaker setting on overall system reliability, when the subsequent speeds of likely service restoration are factored into the choice. A system that separates prematurely because of extremely sensitive breaker response settings may nevertheless experience a much higher overall level of reliability compared to a system with very slow response times. Customers in the second system will experience many fewer, annoying, light-flickering outages, but when a permanent short-circuit occurs, the chances are much greater that severe damage will have been inflicted upon power supply equipment that may, in turn, result in a very lengthy outage until the facilities can be replaced or repaired. By comparison, customers in the first system will experience many more annoying bumps in their computers, but few truly prolonged service interruptions, since even with a solid short-circuit, the line will be interrupted before truly catastrophic damage will have occurred, and therefore the repair time should be much shorter. This tradeoff is exacerbated if the failure affects and damages a transformer like the one in substation S-7 in Figure 1, and that transformer is unique with no spares in inventory (a very costly proposition, since many substation transformers are nearly one of a kind because of non-standardization of power system electrical design across the U.S.).

These sensors and relays are distributed throughout the system, and some are designed to sense voltage, others to detect frequency, as well as those that measure the direction and amount of power flow and line impedance. Maintaining voltage in a close band around its design level is important, because low voltage will tend to increase the current flow which overheats devices and, as an example, can cause motors to burn out. By comparison, too high voltage causes electric arcs between adjacent conductors which again can destroy anything in its path. Over-frequency causes motors and generators to spin too rapidly, and that can lead to the destruction of that equipment through centrifugal force. So all of these relays, one way or another, detect and respond to potential threats to equipment; and it is equipment that they protect directly. Indirectly, people may be helped if they face lower repair costs and more rapid system restoration.

4. Possible Terrorist Assaults

As with the study of most infrastructure systems from the perspective of strategies to withstand and/or rebound from a malevolent human assault, this analysis of the electricity network begins and is greatly informed by existing designs and procedures for dealing with natural events. Thus to understand the consequences of a terrorist assault, we merely

need to substitute a conscious physical assault for a lightning strike or a falling tree limb. And we see that the bulk power system is designed to withstand any single such event and to maintain service to all customers without any interruption; in many case two or three such simultaneous events might be of little notice to virtually all customers. Even if such multiple assaults were to bring down the bulk power system, in most cases the automatic, decentralized protection devices are designed to protect the equipment so that the bulk power system can be restored as rapidly as possible following an outage. And while it is much easier to attack isolated portions of the low voltage electric distribution system, this will in most instances cause only localized harm and discomfort, but not result in the widespread regional blackout about which most people are concerned. That would require an assault on the high-voltage, bulk power system, which because of its inherent redundancy would also require coordinated simultaneous assaults on multiple facilities over a wide area. That type of attack borders on all-out warfare, but if conceivable, it needs to be examined in the context of overall system design philosophy.

And, while it may be much easier for terrorists to destroy key isolated portions of the low-voltage electric distribution system, particularly where facilities are above ground, the resulting service interruptions, while prolonged, would most likely be limited both in their geographic scope and in the number of customers affected. Furthermore, the speed of restoration would be inversely related to the number of simultaneous, geographically related hits, since the limiting factor for restoration is the number of trained line crews that are (and that can be made) available in the area of assault. For this purpose, it would be prudent to station crews in a dispersed geographic pattern.

In large metropolitan areas where the distribution system is usually underground and frequently configured as a network rather than a radial system, the considerations might be more similar to the subsequent analysis of the high-voltage bulk power system. And because redundancy and alternative paths are built into those low-voltage distribution networks, the impact of simultaneous distributed assaults would probably be far less severe than on an overhead radial system. Bringing an entire local network down (there are 21 separate such networks serving NYC) would require a coordinated simultaneous assault on separate feeds (there may be up to twenty such separate feeds into a single NYC network). Usually, these networks are designed to withstand a loss of 20 percent of their feeds at peak load periods, but in fact during the 1999 blackout of the Washington Heights neighborhood in NYC [2] that occurred during a prolonged August heat wave, half of the feeds into that network were lost, yet all customers were still being served, before the operators elected to disconnect the remaining feeds and place the customers out of service. This eventual neighborhood blackout was selected in order to protect the remaining facilities from damage due to overload. Because of this decision, the service was restored within two days, probably far faster than would have been the case had the remaining facilities remained at risk, requiring a much greater number of repairs.

The remaining focus of this analysis, therefore, will be on the tightly coupled high-voltage bulk power system, and how the possible number of assaults might affect the system design and desired interconnectedness. Furthermore, this problem can be thought of in a hierarchical manner. As an example, how many utilities with their own separate

area control centers should be tightly connected with each other and have their operations coordinated and controlled by a power pool operator or an ISO/RTO? The answer hinges in normal times, in part, on how many other neighboring ISOs and internal actions can assault that ISO and affect its stability. At a higher level of analysis, given recent economic pressures plus the fallout from the August 14, 2003 blackout to forge larger coordinated region-wide virtual RTOs by tightly coupling neighboring ISOs, the question remains: how many ISOs should be grouped together, and how big is too big in terms of maintaining system reliability. This question of how large is too large is being asked increasingly about the ever-expanding PJM RTO, even absent consideration of possible terrorist activity. Furthermore, as electricity control areas grow ever larger, is the proper operating and reliability philosophy still to float together or sink together? Under what circumstances should the components be separated so that pieces might be saved in order to reassemble the entire system more rapidly, and how do those guidelines change as systems grow larger?

A strategic response cannot be to rely solely on hardening each of the constituent parts of the system in order to improve the average survivability of the aggregate system, if this is a tightly coupled complex system involving many agents. As an example, Wang and Thorp have shown through many numerical simulations of a bulk power system [3] that the probability of a cascading failure leading to a blackout remains at about once every 35 years, even if the reliability of individual components is improved. It is the degree of interconnectedness of the system that can dominate the expected frequency of catastrophic events and not just the reliability of individual pieces, even though increasing the reliability of those components, including the weakest link, will improve the average reliability of the entire system.

5. Insights through Numerical Simulation

Conceptual insights to possible approaches to these questions can be gleaned from earlier analyses by Levitan, Lobo, Schuler and Kauffman [1] on ways in which organizational performance and stability vary with the size and connectedness of organizations in a stochastic environment. This analysis explores a very simple question: what is the optimal number of similar agents to have working together (behaving under a set of coordinated rules) where each is engaged in the same activity, but when they are jolted periodically by some external force that is not subject to the group's rules? So, the example could be one of how many area control centers are coordinated by an ISO/RTO, or how many ISOs are linked together formally by an over-arching set of coordinating rules? These models presume that the individual operators (agents) are periodically given the opportunity to try to improve their performance (e.g. learning is included explicitly in these simulations), and each agent has the one period foresight to know whether the available change will improve or decrease their individual performance. Therefore, each agent also has the freedom to accept or reject the available change. Alternatively, these simulations can be thought of as a set of trial and error experiments where the random change is forced upon the individual agent (and the group), but they can always return to their prior state if that turns out to have been better.

When several agents are combined in a group, having one participant change their mode of operation affects the performance of all other members of the group. Therefore it is best for the group to agree upon a commonly applied acceptance rule, namely each member of the group will adopt an available change only if it improves the overall group performance (even when the individual performance of the deciding agent might decline). Obviously, in order to make such rules palatable to individual performers, each agent must share a portion of the group's improved performance, even when their own individual contribution declines while that of their colleagues' increases. Otherwise, how could the agent be induced to behave in the interest of the entire group? One simple pay-off rule is to share the group's performance equally, under all circumstances, with all members. In fact, if the performance criterion is the reliability of the bulk power electricity system, this egalitarian "sharing" formula for the group's performance is realistic since all local control areas in a pool experiences a similar level of reliability in terms of avoiding a major blackout. In this example, then, each agent is randomly assigned a sequence of random shocks, at which point the agent must decide to accept or reject the change, according to the group's predetermined criterion. In these generalized simulations, any bias is removed from the choice by also assigning the performance value associated with each change randomly.

Thus depending upon the individual agent's original contribution to group performance, in combination with all group members' performances, a chance to change the way a single agent performs will depend upon not only how that agent's performance would change, but also the effect of that change on how all other members in the group would perform, where these performance values are randomly assigned. Typically, the number of performance states for each agent is limited, and in most of the exercises reviewed here, that number of states will be limited to two (e.g. a switch is either open or closed, a generator is on or off, etc.). Nevertheless, since the performance of each agent hinges on the states of all of the other agents in the group, even with only two members, each with only two states, the entire group has available to it four possible combinations of states, and therefore four potential different payoff values. With only two states, but three participants in the group, the number of possible different performance levels is eight, so where S equals the number of states and L equals group size, the number of possible combinations of states, and therefore of group performance values is equal to S^L .

Therefore the process of finding the highest possible level of performance can be viewed as a random search over a landscape, where the landscape is comprised of all possible combinations of states as vertices. One final determinant of the outcome is the process used for searching for improvements. What has been described previously is the mechanism by which a potential improvement is selected or rejected by an individual participant. Which participant's turn it is to choose is selected randomly, and only one participant gets to choose at a time. This last point is important, because this type of incremental search can result in a group becoming stuck at a local optimum and never reaching its highest possible level of performance; nevertheless, most changes in organizations are incremental, unless subjected to a cataclysmic disruption.

5.1 A Model of Organizational Performance

An example of this search is illustrated in Figure 2 for a group of two participants, each with two possible states. In the example, the binary states are represented by [0;1], and each participant's performance, in combination with its partner's performance, is selected randomly over the unit interval (as an example from the uniform probability distribution). Through a monotonic transformation, this performance level could be transformed into any measure the group valued (e.g. proportion of maximum profit, sales, etc. or "fitness" in a biological context), but since the purpose here is to explore system reliability, the outcomes should be scaled to the unit interval, although perhaps not linearly.

(Insert Figure 2)

In this illustration in Figure 2, suppose we begin in the states [0;1] represented by the bottom box with a total group performance of 0.4 (average per agent of 0.2). Suppose the right hand agent is selected randomly to consider changing her state to a 1, so the group situation is now [0;0] as shown in the left hand box. In this case, the right hand agent would accept this change, not just because her own performance increased from 0.3 to 0.9, but also because the group's total performance improved to 1.2 (a group average of 0.6). If at the next random shock, the right hand agent were again selected randomly to consider a change of state, the only option available to her is to choose state 1, and since her partner is still in state 0, she would have to consider returning to the bottom box. This she will not do because that move represents a reversion to the previous lower level of average group performance of 0.2; therefore, the right hand agent (and the group, by its rules) will choose to remain in the left hand box in state [0;0]. If at some still later time, the left hand agent were randomly selected to consider a change, he would have the opportunity to choose state 1 which would imply the opportunity to consider moving to the situation in the top box. Even though his performance improves from 0.3 to 0.4, the left hand agent will not select this change in state because the right hand agent's performance drops from 0.9 to 0.6, or a decline in total group performance from 1.2 to 1.0 (average performance declines from 0.6 to 0.5). This group has found an optimum location in the left hand box, but it is a local optimum at which it is stuck, since there is a superior performance combination in the box on the right side at states [1;1]!

In fact there are many ways in which this global optimum of 1.6 (average performance of 0.8) might have been found, but it cannot be reached from state [0;0] by the incremental search procedure; both agents would have to change their states simultaneously to discover this option. With the participants restricted to incremental searches, the optimum can only be found by chance. In the Figure 2 example, the initial random assignment of states must be in either the bottom, right or top box (if the agents are assigned initial states [0;0], they will always stay there), and if in the top box, the right hand agent must be selected randomly to choose first (otherwise, if the left hand agent is selected first, he will choose to move to the left, suboptimal box, and the group will be stuck there). Similarly, if the initial state assignment is [0;1] (the bottom box), the left hand agent must be randomly selected to consider the first change: otherwise if the right hand agent is picked first, she will opt to move to the left side box [0;0] which again is suboptimal. Of

course, any group lucky enough to be assigned states [1;1] in the first place will choose to remain there throughout all subsequent trials.

Although it may seem unreasonable for a group's final performance to hinge so crucially on the luck of the initial draw of positions, and also on the subsequent random selection of the sequence in which the agents get to try different states in combination with the others in their group, these simulations are repeated many times (500 times per treatment) with random (and therefore usually different) assignments of initial positions and choice sequences. In this way, it is difficult to attribute any conclusion to the arbitrary initial assignment of states or of sequence of choice. The results that are reported are averaged over these 500 repetitions.

The objective of these simulations is to observe regularities in the relative performance between groups of different sizes. But as the size of each group increases, so does the complexity of possible combinations of states, so another experimental design issue is how long to run the simulation (how many random shocks)? The simulations reported are continued until the group's performance stabilizes at a particular value, usually far less than 2000 iterations in the cases investigated here. Outcomes for group sizes ranging from one to as large as eleven entities are simulated, but over this range, the overall group performance reaches a maximum and stabilizes before the 2000 iterations are completed.

5.2 Modeling the Effect of Externally Transmitted Shocks

An important additional complication, particularly for analyzing the effect on electricity system performance of shocks precipitated by neighboring ISOs or terrorist activity, needs to be included in these simulations. So far, the shocks imposed on the system are completely random, and they might be due to weather and natural events, technological innovation and/or changing customer patterns, but they are not related to the conscious choice by some other group. Therefore, in these simulations that may have anywhere from 9 to 100 other groups acting autonomously, the possibility of having the actions in one group affect the performance of another is added. The variable, J , reflects the number of connections each group has to other external groups (the identities of these connections are selected randomly), and what is different about J , as compared to the relationship among agents within a group, L , is that given a chance to select another state, any agent only considers the effect upon members of its own group, and not upon the J interconnected groups that also will have their performance influenced by her choice. This is an externality in economic terms, and it realistically represents the possible interactions between neighboring ISOs when their activities are not tightly coordinated.

These J relationships are also used to illustrate the possible consequences of terrorist activities; although, in that case the parallel may not be as strong. In many instances agents in the malevolent group will base their selection of an alternative state based upon how well it satisfies their own performance goals, but these are measured in part by how badly it affects the neighboring group. In short, the terrorist may consider inter-group effects explicitly in deciding to accept a possibility. However, in the context of this existing model by Levitan, et. al., since the terrorists' objectives (e.g. headlines, and/or

body count), may not be the same as that of the other groups that represent neighboring power systems (who are trying to improve their own reliability), a decline in one may not be directly proportional to an increase in the malevolent group's satisfaction. Thus the more "neutral" connectivity represented by these simulations may nevertheless be illustrative of the relative effects that these external connections might have on the qualitative relationship between the optimal size of the power system and the number of external influences. The key here is that those external impacts on the power system would fluctuate in response to activity by members of those external groups trying to improve their own performance, in isolation from the objective of the agents in the power system.

5.3 Results of Numerical Simulations

In the first instance, think of a generic organization trying to estimate what its optimal group size should be. Figures 3.a&b illustrate the performances of a variety of group sizes over both time, and for different numbers of connections with external groups. What is true in all of these simulations (there are always 100 agents total in each simulation, so where $L=1$, there are 100 different "groups", and where $L=11$ there are only 9 groups) is that for only a few trials (called "generations" which crudely represents elapsed time if the random shocks hit the agents at a constant rate) larger groups perform better. But as shown in Figure 3a, after approximately 15 periods the performance of smaller groups begins to eclipse that of the larger ones, and in this case where there are no external connections between different groups, by the time the performance stabilizes at about 300 generations, agents acting alone pull ahead of the groups of 11.

(Insert Figs. 3a&b)

What's going on? In this first set of simulations with $J=0$, we can think of each group as searching for the largest of the group's feasible number of "order" statistics. As an example, with $L=1$ and each individual allowed only two states, each group (here an individual agent) is searching for the largest of two order statistics. When the performances are drawn randomly from the uniform density, the expected value in this case is 0.67 , $(n-1)/n$, and Figure 3a shows that after about 300 generations all 100 agents have been selected randomly to try a different state, and they have, on average achieved that expected highest level of performance, 0.67 . But why then doesn't the group of size 11 outperform these individuals acting alone, since the expected value of the largest of 11 order statistics is 0.91 ? Two factors interfere with these larger groups from reaching this isolated ideal. The first cause is a statistical phenomenon. As group size increases, the group payoff becomes the sum of the random draws from a uniform probability distribution, and the probability of that sum is no longer uniform! It becomes a truncated Beta probability distribution, and as an example for groups of two, the sum follows a triangular distribution. In fact as L increases, the central limit theorem applies and the distribution of the payoffs approaches a normal probability distribution. So counteracting the search for an ever-greater available order statistic as group size increases is the fact that the probability mass is concentrating around the mean and away from that highest order statistic.

However, there is a second factor, the search process itself that is muting, on average, the attainment of the highest conceivable performance. As illustrated in Figure 2, there is a possibility for any group larger than one of getting stuck at a local optimum that is less in value than the greatest possible payoff. In fact that possibility of getting stuck at suboptimal levels of performance increases dramatically as group size increases. As an example in Figure 3a, the groups of three ($L=3$) attain the greatest performance, on average, beyond 300 trials. But that value is approximately 0.69 as compared to the theoretical maximum which is the expected value of the largest of 8 order statistics (here = 0.88).

As we add the effect of external shocks in Figure 3.b, complicating and partially offsetting phenomena appear to be arising. First, even with $J=4$ as shown in Figure 3.b, the performance of agents acting alone is much worse than with $J=0$. That is because with four different external agents affecting each group of one, every time an agent finds its higher of two order statistics, another connected group acts. And while the interactive effect is selected randomly, the closer the agent is to its highest order statistic, the greater the probability is that he will be knocked down to a lower value. With sufficient frequency of external shocks, the best this agent acting alone can do is no better than flipping a coin (0.5). Conversely, the larger groups seem to perform relatively better as the number of external interactions increases. This is primarily due to the fact that an external shock may push a group that is stuck at a local optimum into a different set of potential payoffs, so it is free again to search for a greater optimum. As an example, with $J=12$, $L=5$ emerges as the eventual best performer, on average, and the groups of three who were the best long run performers with $J=4$ are now being pushed below the other groups. With $J=18$ as in Figure 4.b, only the larger groups seem able to rebound and withstand the frequent external shocks and still deliver a performance level similar to the outcomes where $J=0$.

(Insert Fig. 4a&b)

So, there seems to be a positive correlation between the optimal group size and the number of inter-group connections; however, inferences also have to be modified by the time frame over which these effects are being analyzed. Figures 4.a&b emphasize how a fairly sharply peaked preference for a particular group size emerges after 200 or more trials (generations), but when the concern is to sustain performance on average from the outset, then the preference is always for very large groups (here 11). And if there is a choice of both group size and the number of internal connections, then as suggested by Figure 4.a, group sizes no larger than three with four external connections might be optimal. In the case of ISO design, that might suggest that the NYISO with four semi-autonomous external connections (NEISO, PJM, IMO and Hydro-Quebec) should probably span no more than three area control centers within the NYISO. In fact, the NYISO includes three large utilities, but it also has four other smaller entities plus the New York Power Authority (NYPA). Of course, particularly with its recent expansion to include AEP and the MISO, the PJM RTO encompasses a far greater number of area control centers, but as its geographic scope increases, so too does the number of external connections.

Figure 5 illustrates how for longer time horizons (generations in excess of 100), a ridge of optimal performance begins to emerge suggesting that as J increase, so should L , the desired group size. This suggests that in response to increased dispersed terrorist activity (greater J ?) recovery and overall performance might be greater with ever-larger-sized coordinated units. But Figure 5 also suggests there is a price to be paid by getting bigger and more coordinated, not only in terms of those costs of coordination, but also in terms of reduced performance because of performance failures, particularly in the long run. As emphasized in Figure 5, once a system has attained the optimal group size for any particular J , increasing L even further results in an average loss of performance. Nevertheless, the general rule is the larger the number of external effects that are expected, the larger the number of members that should be included in the internal group.

(Insert Fig. 5)

Not only is the average performance of each group affected by the combination of external connections and group size, so too is the stability of the outcome, as illustrated in Figure 6. This diagram shows one measure of the extent to which each agent is selecting new states after a particular period of time has elapsed. As an example, Figure 6 shows that even after 2000 generations, agents acting alone ($L=1$) but with 5 to 10 external connections, are still moving to other states in more than 40 percent of the randomly selected opportunities to choose (in this case generated mostly by changes in states by externally connected agents). With $L=1$, every time an agent finds the largest of their two order statistics, a change in state by some externally connected agent alters the performance values of all states for agents with which they are linked, thereby setting in motion an additional set of possible choices.

(Insert Fig. 6)

A consistent pattern that seems to emerge by comparing Figures 5 and 6 for identical durations (generations) is that there is also a positive relationship between group size, L , and the number of external connections per group, J , that leads to higher flip rates, and therefore greater instability, that parallels the combinations that lead to optimal performance (e.g. system reliability). As an example, in Figure 6, for any particular J , the flip rates settle down to extremely low levels by 2000 generations for a sufficiently high L (large group size). As, however, the number of external connections continue to increase for any given size group, the flip rates begin to increase rapidly indicating a region of instability. Furthermore, in comparing the ridge of optimal performance in Figure 5 with the regions of instability in Figure 6, those flip rates begin to rise as the number of external connections rises above the level that yields the best performance. The message seems to be that optimal system performance borders a region of instability! So not only is it desirable to get the right group size in relation to the anticipated number of externally imposed shocks in order to maximize the system reliability, it is extremely important to err on the side of having groups that are larger than optimal in order to avoid being driven into a region of instability should the frequency of those external shocks increase. In this instance, that unstable region with higher flip rates might be representative of more frequent extensive blackouts, even though the average performance level does not fall precipitously until J becomes significantly larger, because

the higher flip rate implies groups being knocked off their locally optimal states more frequently and therefore searching continuously for better outcomes.

6. Potential Inferences

Although this discussion emphasizes the robustness of the high-voltage bulk power system in the U.S., precisely because of its decentralized nature, some principles to guide improvement do emerge. First, because of the network configuration of the transmission grid, plus the settings on the automatic decentralized protection equipment that are biased to preserve the hardware, even if an extensive blackout were to occur, many areas may be restored to service quite rapidly, usually within a day or two. In order for an assault on the bulk power system to result in a prolonged region-wide outage, a massive region-wide coordinated attack on multiple key facilities would be required. However in that instance, one aspect of the “hodge-podge” nature of multiple, non-standardized, ill-coordinated electric systems could create difficulties. To the extent that key equipment is not standardized electrically, (e.g voltage, phase configuration, etc.) across control areas or individual utilities, spare equipment may not be available in inventory. To the further extent that an increasing portion of that equipment is being manufactured abroad, the lag in securing replacements may be prolonged (or worse, if under the control of an ally of the inflictor of the original trauma) with attendant delays in restoring total service to everyone. However, most surrounding regions would be patched back into service rapidly, awaiting the replacement of the damaged equipment. In all cases, the potential for long-lasting trauma to the U.S. would also depend on the extent of collateral damage that resulted from the original power outage, and that in turn would hinge upon the timing (associated weather) or social conditions that were coincident with the triggering electrical event. One example would be a prolonged power outage in the north during an extended period of subfreezing weather. In that event, many automated space heating systems would be interrupted, risking the freezing and bursting of water pipes, and thereby causing enormous widespread distributed damage to many customers.

In the case of the power grid, therefore, while a “signature” event capable of garnering widespread publicity, like a major regional blackout, is possible, the actual damage is likely to be relatively small in most instances. The regional blackout is a failure of the bulk-power system that because of redundancies can be re-established rapidly in most instances. By comparison, it is widespread failures of the low-voltage distribution system that leads to prolonged (up to a week or more) outages of many customers, primarily because repair crews must attend to nearly every customer before all service is restored. Evidence of this comparison in terms of natural events is the northeast blackout of August 14, 2003 (a bulk power system failure) that is estimated to have cost society \$7 billion but where most service was restored by the next day, as compared to the consequences of the hurricane season in Florida this past August and September 2004 (largely widespread failures of the low-voltage distribution system) in which many customers were out of service more than a week at estimated costs that are still being assembled but may be many times greater.

So, with respect to designing and coordinating the bulk power system, in addition to the question posed at the beginning - - “Float together or sink together?” - - the question must

be added that if they sink, how rapidly can the life boats be re-floated or replaced? Current operating practice seems to err on the side of ensuring rapid restoration by shutting off early and protecting equipment against catastrophic damage. The remaining set of questions relates to the scope of coordination among operating entities.

How many utilities and/or area control centers should be coordinated through a single ISO, and how many neighboring ISOs should be combined into a single virtual RTO? In using the numerical analysis of Levitan, et. al. [1] the answer in part depends upon how tightly coupled the agents and group become. If individual agents, or groups acting as a single agent, are each subject to individual choices (shocks) but agree to select alternatives based upon the collective, rather than individual good, then as the number of connections to external institutions or activities increase, so too should the number of cooperating entities increase. This increased group size with increased external interactions should not only improve average group performance above what it might otherwise be, it should also reduce the chances of non-stable behavior in the face of shocks. By comparison, if the agents within a group are so tightly coupled that a shock to one seems like a shock to all, and the response by one is a response by all, then in terms of this analytic paradigm, all such tightly coupled groups are merely behaving like individual agents, and the proper analogy is the case where $L=1$. In virtually all cases, this is shown to yield inferior performance, both in terms of learning, average reliability and stability of results, particularly as the number of connections with other external groups increases.

In the cases of loosely coupled but coordinated groups, the general guideline is to have the size of these groups whose choices of state are determined by overall group performance, not just by that of individual members, increase as the degree of external connectedness increases. A somewhat different relationship between optimal L in response to the anticipated J is obtained depending upon over what duration of shocks the performance is to be gauged. After 2000 generations, the number of random shocks experienced by each group will have increased twenty times on average, as compared to 100 generations. In each case, this number of cumulative shocks experienced, and therefore the number of state changes evaluated, will increase also as the number of external connections, J , increases. This offers one heuristic explanation of why, for optimal average performance, group size should increase as J increases; the number of random choices to be evaluated within the group keeps pace with the frequency of externally inflicted possible state changes.

But how does this relate to optimal configurations of bulk power control areas for electricity supply? We have previously suggested that for dealing with random natural shocks, where $0 \leq J \leq 5$, $L=3$ is the optimal group size. This implies that the NYISO, as an example, with four external ISOs on its borders capable of transmitting shocks, should be comprised of at least three semi-autonomous internal area control centers, each of which acts, however, according to some well-agreed-upon group enhancing criterion. In fact, the NYISO consists of seven utilities plus the New York Power Authority, perhaps too many to achieve optimal average performance, but sufficiently large to withstand additional external shocks (added J) from malevolent agents. And the slightly larger than

optimal L does provide some assurance of not drifting into the unstable regions highlighted in Figures 6.

Similarly, early proposals to group three of the first ISOs (NEISO, NYISO and PJM) together into a large regional RTO, satisfied this internal group size criterion of $L=3$, except as the geographic area covered by the coordinated unit increases, so too does the number of interconnected but non-coordinated groups on the border (in this case 9 or 10), so that the optimum group size should also be increased (which might further increase the size of J). So if grouping is done by contiguous units geographically, a key concern about enlarging L is that J may grow by a larger proportion, thus increasing the chances that the group will be driven into an unstable region. It is that region of high flip rate that indicates more frequent collapses into low performance states following an external shock that might be representative of a major blackout.

Conversely, a propensity to separate into semi-autonomous, self-sufficient units in the face of external shocks may reduce the extent of shock transmission through cooperating units, but again it is the ratio of J/L that is important, and an examination of Figures 5 and 6 shows that for $L=1$, fairly high levels of average reliability are available so long as J remains very small, but if the frequency of assaults increases even marginally, the performance drops precipitously and the instability soars.

Before concluding precisely what the optimal set of connections should be of neighboring power control areas, however, refined calibrations and some re-arrangement of interconnections, in these illustrative simulations, need to be performed. In particular, the impact of a shock from an externally connected group is drawn randomly in these simulations. That means the consequence could be positive as well as the negative effect desired by a terrorist, as an example (note, however, that in the simulations described here, as average group performance increases with increased iterations (generations), the results of a randomly drawn new payoff following an external shock are more likely to be smaller rather than larger). Another realistic modification would be to explore the effects of smaller group sizes for malevolent actors, as compared to the electricity control units; in these simulations, all groups are of the same size.

Nevertheless, several general observations can be drawn from this discussion. First, while it pays to have group size increase as the frequency and number of sources of potential externally imposed shocks increases, it pays to do so in a loosely coupled way where each shock hits a sub-set of the overall group, even though the response is coordinated for the average benefit of the entire group, and not for the initially impacted sub-group alone. Second, it may be all right to risk the collapse of the entire system, if in doing so, equipment is spared so the restoration can be rapid. And third, “hodge-podge” can be beneficial, so long as there is an agreed upon, coordinating objective.

References

1. Levitan, B., Lobo, J., Schuler, R., & Kauffman, S., "Evolution of Organizational Performance and Stability in a Stochastic Environment", *Computational & Mathematical Organization Theory*, 8, 2002, pp.281-313.
2. U.S. Department of Energy, "Report of the U.S. Department of Energy's Power Outage Study Team", Washington, D.C., Interim Report, January 2000 & Final Report, March 2000.
3. Wang, H. & Thorp, J., "Optimal Locations for Protection System Enhancement: A Simulation of Cascading Outages," *IEEE Transactions on Power Delivery*, 16, 4, October 2001, pp.528-533.

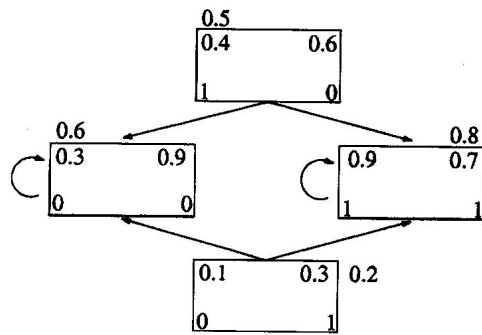
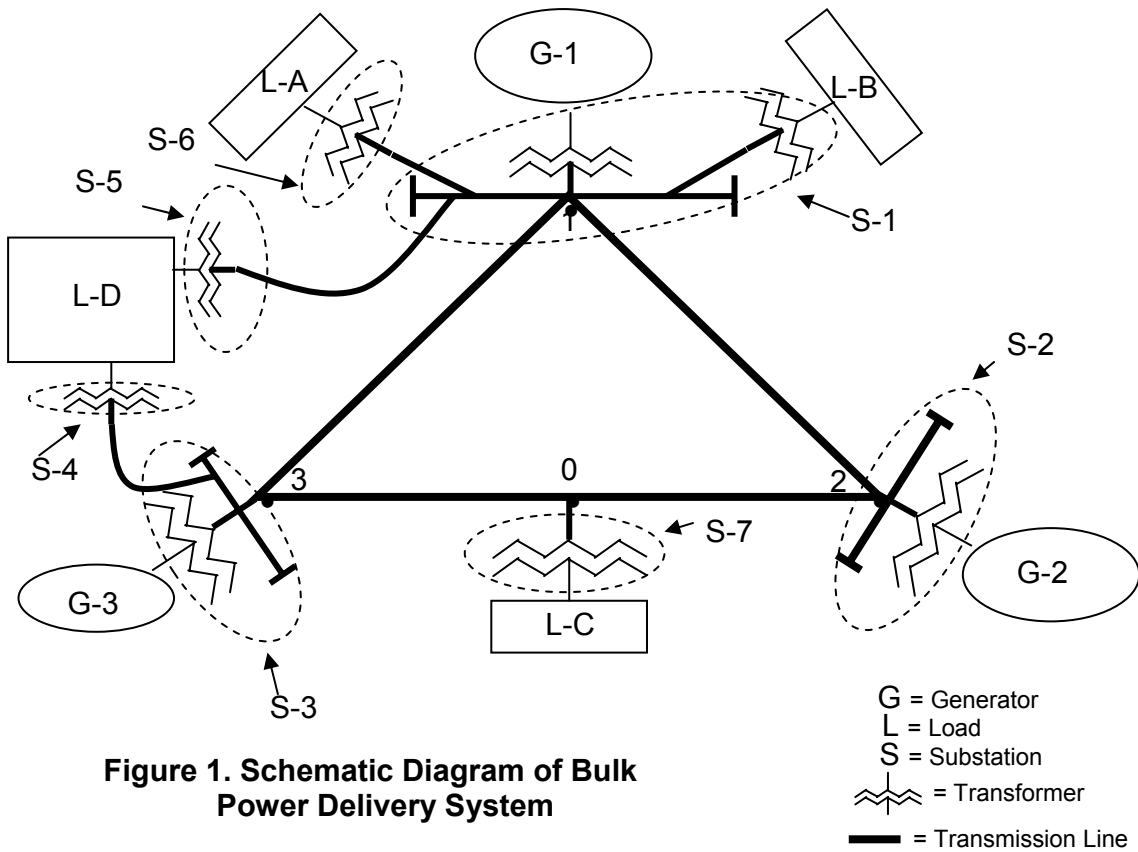


Figure 2. Possible payoffs for a group of size $L = 2$ with $S = 2$ and $J = 0$. The two columns of decimal numbers inside each box denote the payoffs to agents 1 and 2 under the given configuration; the decimal number outside each box denotes the group's average payoff. The arrows indicate state transitions that increase average group payoff.

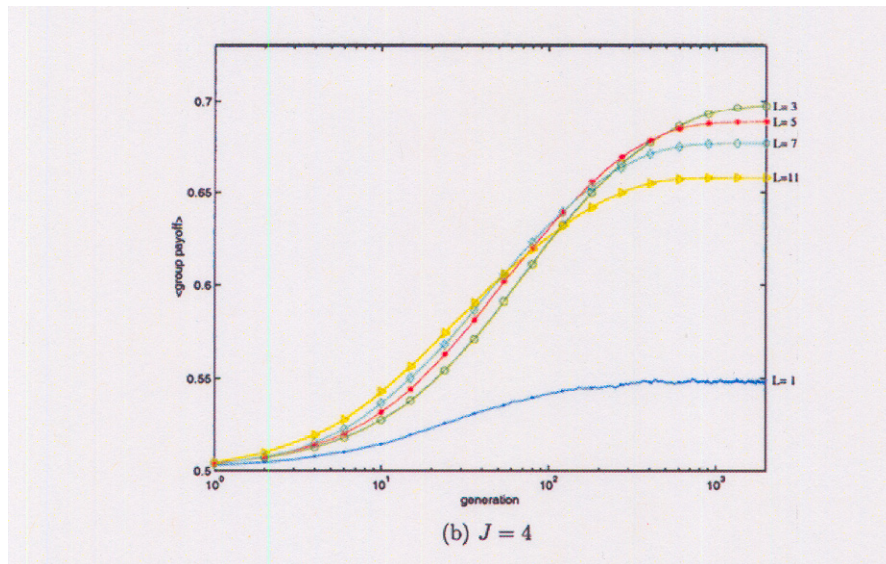
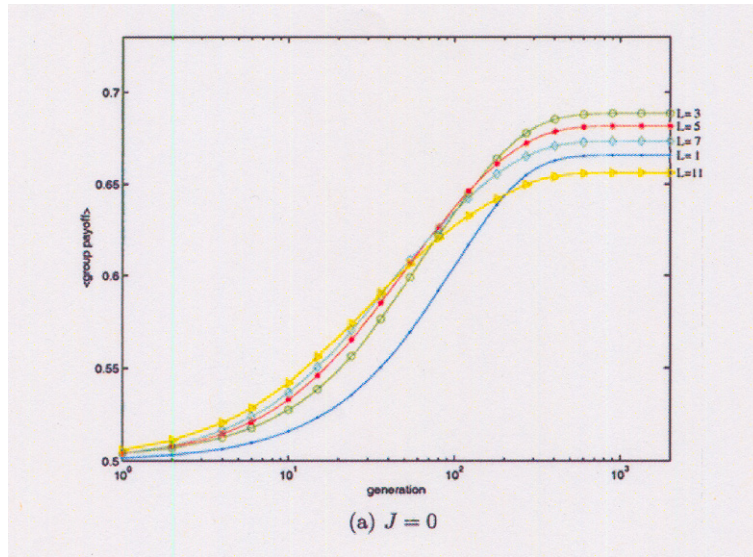
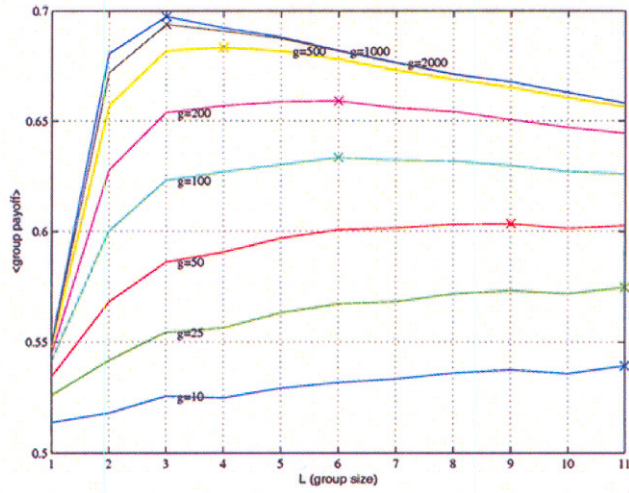
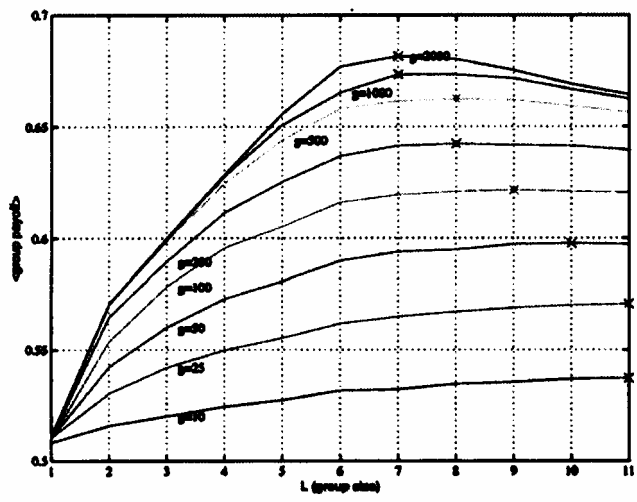


Figure 3. Average group payoff over time (generations) for different group sizes, L , and two different levels of externalities, $J=0$ and 4 .



(a) $J = 4$



(b) $J = 18$

Figure 4. Average group payoff per group size (L) for varying generations (g) and two different levels of externalities $J = 4$ and 18.

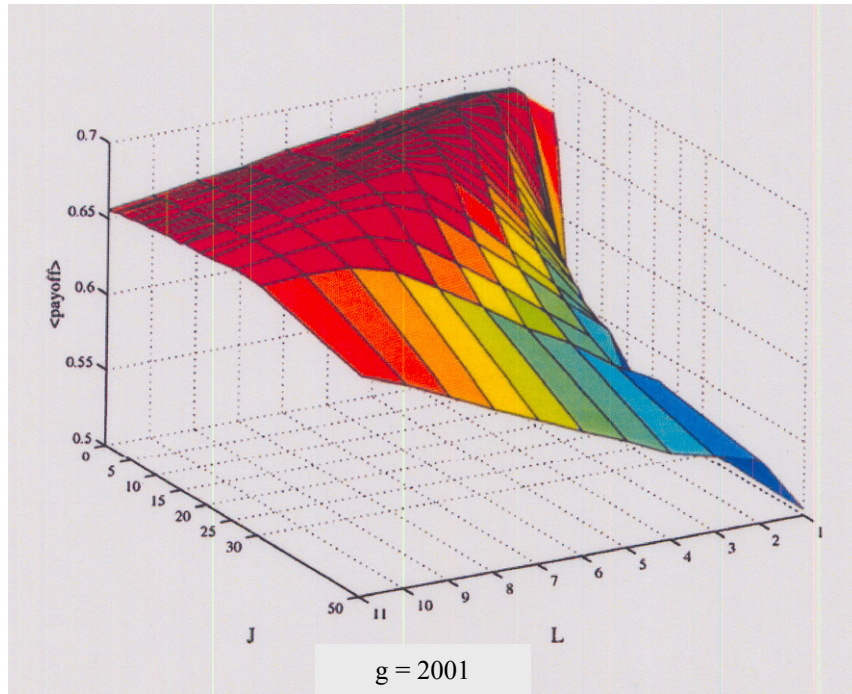


Figure 5. Average group payoff as a function of group size, L , and level of externalities, J , after 2001 generations.

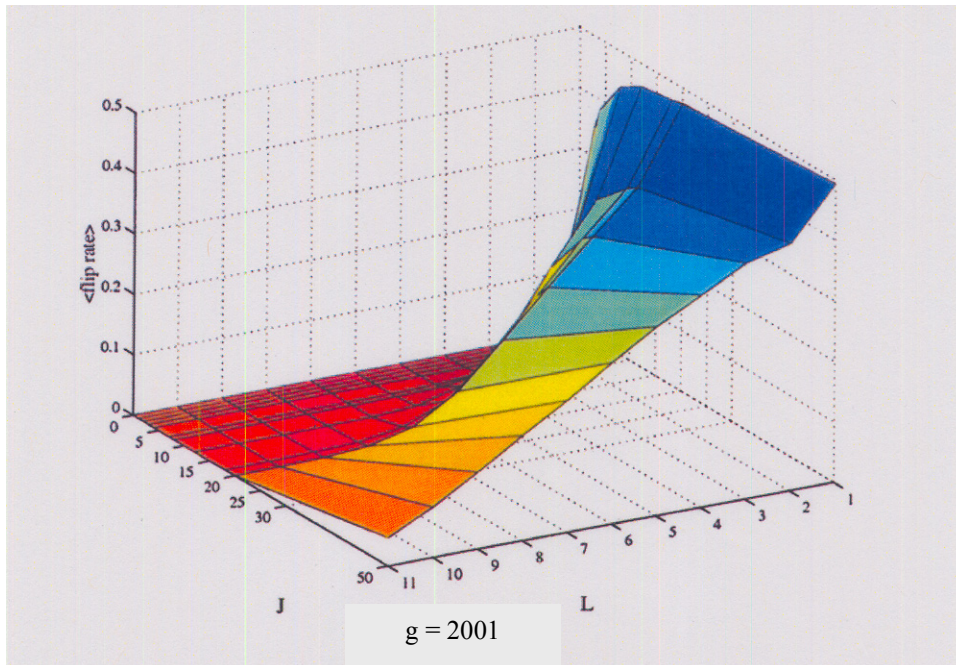


Figure 6. Average flip rate as a function of group size, L , and magnitude of externalities, J , for 2001 generations. (See main text for definition of “flip rate”).